

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 17:15:53 ; Search time 5490.32 Seconds
(without alignments)
9970.316 Million cell updates/sec

Title: US-10-677-877A-1
Perfect score: 963
Sequence: 1 aagcttacgtaagatctaac.....gtaaacctccctcatctaga 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*
1: gb_ba.*
2: gb_in.*
3: gb_env.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pr.*
9: gb_ro.*
10: gb_ats.*
11: gb_sy.*
12: gb_un.*
13: gb_vi.*
14: gb_htg.*
15: gb_pl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	942.2	97.8	5921	6	CQ833976 Sequence
2	942.2	97.8	5921	6	CS063169 Sequence
3	942.2	97.8	5921	6	CS080950 Sequence
4	942.2	97.8	6728	6	AX167114 Sequence
5	942.2	97.8	6728	6	AX330012 Sequence
6	942.2	97.8	6728	6	AX330557 Sequence
7	942.2	97.8	6728	6	AX333300 Sequence
8	942.2	97.8	6728	6	AX333330 Sequence
9	942.2	97.8	6728	6	AX411302 Sequence
10	942.2	97.8	6728	6	HSPPA11CO
11	937.4	97.3	4721	8	AB209597 Homo sapi
12	937.4	97.3	3347	6	CQ724221 Sequence
13	935.4	97.1	3347	8	AX776246 Sequence
14	935.4	97.1	3347	8	HUMCG1PAL
15	934.2	97.0	4752	8	BC036531 Homo sapi
16	930	96.6	4409	6	A91174 Sequence 1
17	930	96.6	4409	6	BD070009 Recombina
18	930	96.6	4409	6	AR397575 Sequence

19	862.2	89.5	4748	6	AX146420
20	850.4	88.3	4425	6	AX146426
21	843.2	87.6	4696	4	AF153062
22	826.2	85.8	855	8	S64596
23	801.4	83.2	5721	9	RNCAIT1
24	790.2	82.1	2013	9	BC003198
25	790.2	82.1	4038	9	BC059281
26	790.2	82.1	4711	9	BC050014
27	780.6	81.1	4589	9	MMU08020
28	778.6	80.9	1205	9	MNCIAI
29	778	80.8	778	6	AX806548
30	708	73.5	1552	5	GGCOL9
31	676.4	70.2	678	8	HSCOLA1P
32	664	69.0	664	6	AX806633
33	654.8	68.0	5711	5	BC082718
34	654.2	67.9	659	6	AX193424
35	651.6	67.7	5718	5	BC049829
36	645.4	67.0	6272	5	AB015438
37	638.2	66.3	2625	5	AB008373
38	638.2	66.3	4651	5	AB052835
39	636.2	66.1	642	6	AX193399
40	634.2	65.9	639	6	AX193436
41	632.2	65.6	637	6	AX193451
42	631.6	65.6	650	6	AX193400
43	631.2	65.5	636	6	AX193438
44	631.2	65.5	636	6	AX193455
45	625.6	65.0	5650	5	AB015440

ALIGNMENTS

RESULT 1	CQ833976	Sequence 12 from Patent EP1439393.	5921 bp	DNA	linear	PAT 29-JUL-2004
LOCUS	CQ833976					
DEFINITION	CQ833976					
ACCESSION	CQ833976.1	GI:50833581				
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
ORIGIN						
Query Match						
Best Local Similarity						
Matches						
Qy	13	GATCTAACGGTCTCCCTGGGCCCATTTGGGCCCTCGTCTCGGGTGGCACTGGTGATG	72			
Db	3580	GACTCAAGGTCTCTCCCTGGGCCCATTTGGGCCCTCGTCTCGGGTGGCACTGGTGATG	3639			
Qy	73	CTGGTCTGTGTGTGTCCTCCCGGCCCTCTGTGACTCTCTGTGTCCTCCAGAGGCTCAGATGGTGGCC	132			
Db	3640	CTGGTCTGTGTGTGTCCTCCCGGCCCTCTGTGACTCTCTGTGTCCTCCAGAGGCTCAGATGGTGGCC	3699			
Qy	133	CTGGTTCGACTTCAGCTTCTCTGCCCCCAGCCACTCAGAGAGGCTCAGATGGTGGCC	192			

[illegible]

KRHVFEESMTDFQFEYGGSDPADVAIQLFRLMSTSEASQNIYHCKNSVAYMD
QOTENLKALLKNEIBIRAEGRNFRFTVSDGTCTSHGTGAWKTVIEYKTKSSR
LPIIDVAPLDVGAPDQBFQDFGVPCFL"
120...185

sig_peptide
ORIGIN

Query Match 97.8%; Score 942.2; DB 6; Length 6728;
Best Local Similarity 99.7%; Pred. No. 5.1e-138;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 13 GATCTAACGGTCTCCCTGGGCCCAATTGGGGCCCTCTGGTCCCTCGGGTCCGACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGGCCCAATTGGGGCCCTCTGGTCCCTCGGGTCCGACTGGTGATG 3639
Qy 73 CTGGTCTCTGTTGGTCCCGGCCCTCTCGACCTCTGGTCCCTGGTCCCTGGTCCCTCCAGCG 132
Db 3640 CTGGTCTCTGTTGGTCCCGGCCCTCTCGACCTCTGGTCCCTGGTCCCTGGTCCCTCCAGCG 3699
Qy 133 CTGGTCTCTGACTTCAGCTTCTCTGCCCGACCTCTCAAGAGAAAGGCTCACGATGGTGCC 192
Db 3700 CTGGTCTCTGACTTCAGCTTCTCTGCCCGACCTCTCAAGAGAAAGGCTCACGATGGTGCC 3759
Qy 193 GCTACTACCGGGCTGATGATGCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 252
Db 3760 GCTACTACCGGGCTGATGATGCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 3819
Qy 253 CCCTCAAGAGCCTGAGCCAGAGATCGAGAACATCCGGAGCCAGAGGGAAGCCGCAAGA 312
Db 3820 CCCTCAAGAGCCTGAGCCAGAGATCGAGAACATCCGGAGCCAGAGGGAAGCCGCAAGA 3879
Qy 313 ACCCGCCCGCACCTGCGTGAATCTCAAGATGTGCCACTCTGATGGAAGAGTGGAAGT 372
Db 3880 ACCCGCCCGCACCTGCGTGAATCTCAAGATGTGCCACTCTGATGGAAGAGTGGAAGT 3939
Qy 373 ACTGGATGTACCCCAACAAAGGCTGAACTCTGATGCAATCAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATGTACCCCAACAAAGGCTGAACTCTGATGCAATCAAGTCTTCTGCAACATGG 3999
Qy 433 AGACTGTGAGACTGCGTGTACCCCACTCAGCCAGCTGAGTGGCCAGAGAACTGGTACA 492
Db 4000 AGACTGTGAGACTGCGTGTACCCCACTCAGCCAGCTGAGTGGCCAGAGAACTGGTACA 4059
Qy 493 TCAGCAAGAACCCCAAGAGGCAATGTCTGGTTCGGCGAGAGCATCACCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGAGGCAATGTCTGGTTCGGCGAGAGCATCACCGATGGAT 4119
Qy 553 TCAGTTCGAGATGCGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612
Db 4120 TCAGTTCGAGATGCGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 4179
Qy 613 TCCTGCGCCTGATGTCCACCGAGGCCCTCCAGAACATCCTACCTGCAAGAACAGCG 672
Db 4180 TCCTGCGCCTGATGTCCACCGAGGCCCTCCAGAACATCCTACCTGCAAGAACAGCG 4239
Qy 673 TGGCTTACATGGAACGACAGACTGGCAACCTCAAGAAAGGCCCTGCTCTCAAGGGCTCCA 732
Db 4240 TGGCTTACATGGAACGACAGACTGGCAACCTCAAGAAAGGCCCTGCTCTCAAGGGCTCCA 4299
Qy 733 ACAGATCGAGATCCGCGCGAGGCAACAGCGCTTCACTACAGGTCATCTGTGATG 792
Db 4300 ACAGATCGAGATCCGCGCGAGGCAACAGCGCTTCACTACAGGTCATCTGTGATG 4359
Qy 793 GCTGACAGGTTCACCGGAGCCTGGGGCAAGACAGTGTGATGATACAAACACCAAGT 852
Db 4360 GCTGACAGGTTCACCGGAGCCTGGGGCAAGACAGTGTGATGATACAAACACCAAGT 4419
Qy 853 CCTCCCGCTGCCCATCATCGATGTGGCCCTCTTGGACGTTGTGGTCCCGAGACAGGAAT 912
Db 4420 CCTCCCGCTGCCCATCATCGATGTGGCCCTCTTGGACGTTGTGGTCCCGAGACAGGAAT 4479
Qy 913 TCGGCTTCGAGCTTGGGCCCTGTCTGCTTCTCTGTAACCTCCCTCCATC 959
Db 4480 TCGGCTTCGAGCTTGGGCCCTGTCTGCTTCTCTGTAACCTCCCTCCATC 4526

RESULT 5
AX330012

LOCUS 6728 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 521 from Patent WO0194629.
ACCESSION AX330012

VERSION AX330012.1 GI:18102990
KEYWORDS

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1

AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,

Horrihan, S., Soppet, D.R. and Weaver, Z.

TITLE Cancer gene determination and therapeutic screening using signature

JOURNAL Gene sets

PATENT: WO 0194629-A 521 13-DEC-2001;

Avalon Pharmaceuticals (US)

FEATURES Location/Qualifiers

source 1..6728

/organism="Homo sapiens"

/mol_type="unassigned DNA"

/db_xref="taxon:9606"

ORIGIN

Query Match 97.8%; Score 942.2; DB 6; Length 6728;
Best Local Similarity 99.7%; Pred. No. 5.1e-138;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 13 GATCTAACGGTCTCCCTGGGCCCAATTGGGGCCCTCTGGTCCCTCGGGTCCGACTGGTGATG 72

Db 3580 GACTCAACGGTCTCCCTGGGCCCAATTGGGGCCCTCTGGTCCCTCGGGTCCGACTGGTGATG 3639

Qy 73 CTGGTCTCTGTTGGTCCCGGCCCTCTCGACCTCTGGTCCCTGGTCCCTGGTCCCTCCAGCG 132

Db 3640 CTGGTCTCTGTTGGTCCCGGCCCTCTCGACCTCTGGTCCCTGGTCCCTGGTCCCTCCAGCG 3699

Qy 133 CTGGTCTCTGACTTCAGCTTCTCTGCCCGACCTCTCAAGAGAAAGGCTCACGATGGTGCC 192

Db 3700 CTGGTCTCTGACTTCAGCTTCTCTGCCCGACCTCTCAAGAGAAAGGCTCACGATGGTGCC 3759

Qy 193 GCTACTACCGGGCTGATGATGCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 252

Db 3760 GCTACTACCGGGCTGATGATGCGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 3819

Qy 253 CCCTCAAGAGCCTGAGCCAGAGATCGAGAACATCCGGAGCCAGAGGGAAGCCGCAAGA 312

Db 3820 CCCTCAAGAGCCTGAGCCAGAGATCGAGAACATCCGGAGCCAGAGGGAAGCCGCAAGA 3879

Qy 313 ACCCGCCCGCACCTGCGTGAATCTCAAGATGTGCCACTCTGATGGAAGAGTGGAAGT 372

Db 3880 ACCCGCCCGCACCTGCGTGAATCTCAAGATGTGCCACTCTGATGGAAGAGTGGAAGT 3939

Qy 373 ACTGGATGTACCCCAACAAAGGCTGAACTCTGATGCAATCAAGTCTTCTGCAACATGG 432

Db 3940 ACTGGATGTACCCCAACAAAGGCTGAACTCTGATGCAATCAAGTCTTCTGCAACATGG 3999

Qy 433 AGACTGTGAGACTGCGTGTACCCCACTCAGCCAGCTGAGTGGCCAGAGAACTGGTACA 492

Db 4000 AGACTGTGAGACTGCGTGTACCCCACTCAGCCAGCTGAGTGGCCAGAGAACTGGTACA 4059

Qy 493 TCAGCAAGAACCCCAAGAGGCAATGTCTGGTTCGGCGAGAGCATCACCGATGGAT 552

Db 4060 TCAGCAAGAACCCCAAGAGGCAATGTCTGGTTCGGCGAGAGCATCACCGATGGAT 4119

Qy 553 TCAGTTCGAGATGCGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612

Db 4120 TCAGTTCGAGATGCGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 4179

Qy 613 TCCTGCGCCTGATGTCCACCGAGGCCCTCCAGAACATCCTACCTGCAAGAACAGCG 672


```
Db 4180 TCCTGGCCCTGATGTGTCCACCGAGGCTCCCAAGACATCACTTACCACTGCAAGAACAGCG 4239
Qy 673 TGGCTTATACGACAGACAGACTGGGAACTTCAAGAGGCCCTGCTCTCAAGGGCTCCA 732
Db 4240 TGGCTTATACGACAGACAGACTGGGAACTTCAAGAGGCCCTGCTCTCAAGGGCTCCA 4299
Qy 733 ACGAGATCGAGATCCGGCCGAGGCGCAACAGCCGCTTCACTTACAGAGGTCACTGTCGATG 792
Db 4300 ACGAGATCGAGATCCGGCCGAGGCGCAACAGCCGCTTCACTTACAGAGGTCACTGTCGATG 4359
Qy 793 GCTCAGCAGTTCACACCGAGCCCTGGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGT 852
Db 4360 GCTCAGCAGTTCACACCGAGCCCTGGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGT 4419
Qy 853 CCTCCGCGCTCCCATCATCATGATGTGGCCCTTGGAGCGTTGGTGGCCCGAGACAGGAAT 912
Db 4420 CCTCCGCGCTCCCATCATCATGATGTGGCCCTTGGAGCGTTGGTGGCCCGAGACAGGAAT 4479
Qy 913 TCGGCTTCGAGTGTGGCCCTGCTGCTTCCGTAAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTGTGGCCCTGCTGCTTCCGTAAACTCCCTCCATC 4526

RESULT 6
AX330557
LOCUS AX330557 6728 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 1066 from Patent WO0194629.
ACCESSION AX330557
VERSION AX330557.1 GI:18103535
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
JOURNAL Patent: WO 0194629-A 1066 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
Location/Qualifiers
1..6728
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 97.8%; Score 942.2; DB 6; Length 6728;
Best Local Similarity 99.7%; Pred. No. 5.1e-138;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 13 GATCTACGGTCTCCCTGGGCCGCAATGGGCCCTTGGTCTCTCGGGTGGCACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGGCCCAATGGGCCCTTGGTCTCTCGGGTGGCACTGGTGATG 3639
Qy 73 CTGGTCTCTGTGGTCCCGGGCCCTCTGAGCTCTGTTGTCCTGGTCCCTGGTCCCAAGCG 132
Db 3640 CTGGTCTCTGTGGTCCCGGGCCCTCTGAGCTCTGTTGTCCTGGTCCCTGGTCCCAAGCG 3699
Qy 133 CTGGTTTCGACTTCAGCTTCTGCCCCAGCCACTCAAGAGAGGGCTCACGATGGTGGCC 192
Db 3700 CTGGTTTCGACTTCAGCTTCTGCCCCAGCCACTCAAGAGAGGGCTCACGATGGTGGCC 3759
Qy 193 GCTACTACGGGGTGTGATGTCGAATGTGGTTCGTGACCGTGACCTCGAGGTGACACCA 252
Db 3760 GCTACTACGGGGTGTGATGTCGAATGTGGTTCGTGACCGTGACCTCGAGGTGACACCA 3819
Qy 253 CCCTCAAGAGCCTCAGCGCCAGCAGATCGAGACATCCCGAGCCAGAGGAGCGGCAAGA 312
Db 3820 CCCTCAAGAGCCTCAGCGCCAGCAGATCGAGACATCCCGAGCCAGAGGAGCGGCAAGA 3879
```

```
Qy 313 ACCCGCCCGCAGCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGT 372
Db 3880 ACCCGCCCGCAGCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGGTGCACCTCGATGCCATCAAAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGGTGCACCTCGATGCCATCAAAAGTCTTCTGCAACATGG 3999
Qy 433 AGACTGTGTAGACTGCTGGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAACTGGTACA 492
Db 4000 AGACTGTGTAGACTGCTGGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAACTGGTACA 4059
Qy 493 TCAGCAGAACCCCAAGGACAAAGAGCATGCTGTGTTCCGGGAGAGCATGACCGATGGAT 552
Db 4060 TCAGCAGAACCCCAAGGACAAAGAGCATGCTGTGTTCCGGGAGAGCATGACCGATGGAT 4119
Qy 553 TCAGTTCGAGTATGGGGCCAGGGCTCCGACCTGCCCAGTGTGGCCCATCCAGCTGACCT 612
Db 4120 TCAGTTCGAGTATGGGGCCAGGGCTCCGACCTGCCCAGTGTGGCCCATCCAGCTGACCT 4179
Qy 613 TCTCGCCCTGATGTCCACCGAGGCTTCCAGAACATCACTACCACTGCAAGAACAGCG 672
Db 4180 TCTCGCCCTGATGTCCACCGAGGCTTCCAGAACATCACTACCACTGCAAGAACAGCG 4239
Qy 673 TGGCTTACATGGACAGCAGACTGGGCAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCA 732
Db 4240 TGGCTTACATGGACAGCAGACTGGGCAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCA 4299
Qy 733 ACGAGATCGAGATCCGGCCGAGGCGCAACAGCGCTTCACTACAGCGTCACTGTCGATG 792
Db 4300 ACGAGATCGAGATCCGGCCGAGGCGCAACAGCGCTTCACTACAGCGTCACTGTCGATG 4359
Qy 793 GCTGCGAGTTCACACCGGAGCTGGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGT 852
Db 4360 GCTGCGAGTTCACACCGGAGCTGGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGT 4419
Qy 853 CCTCCGCGCTCCCATCATCATGATGTGGCCCTTGGAGCGTTGGTGGCCCGAGACAGGAAT 912
Db 4420 CCTCCGCGCTCCCATCATCATGATGTGGCCCTTGGAGCGTTGGTGGCCCGAGACAGGAAT 4479
Qy 913 TCGGCTTCGAGTGTGGCCCTGCTGCTTCCGTAAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTGTGGCCCTGCTGCTTCCGTAAACTCCCTCCATC 4526

RESULT 7
AX333300
LOCUS AX333300 6728 bp DNA linear PAT 09-JAN-2002
DEFINITION Sequence 3809 from Patent WO0194629.
ACCESSION AX333300
VERSION AX333300.1 GI:18123934
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
JOURNAL Patent: WO 0194629-A 3809 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source
Location/Qualifiers
1..6728
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 97.8%; Score 942.2; DB 6; Length 6728;
```

Best Local Similarity 99.7%; Pred. No. 5.1e-138;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 13 GATCTAAGGTCCTCCCTGGGCCCCATTGGGCCCCCTGGTCTCGGGTTCGCACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGGCCCCATTGGGCCCCCTGGTCTCGGGTTCGCACTGGTGATG 3639

Qy 73 CTGGTCTCTGTTGGTCCCTCCGGCCCTCTCGGACCTCTCGGTCCTCCCTGGTCTCCCTCCAGCG 132
Db 3640 CTGGTCTCTGTTGGTCCCTCCGGCCCTCTCGGACCTCTCGGTCCTCCCTGGTCTCCCTCCAGCG 3699

Qy 133 CTGGTCTCTGATTCAGCTTCTCTGCCCCAGCCACTCAAGAGAGGCTCAGAGTGGTGGCC 192
Db 3700 CTGGTCTCTGATTCAGCTTCTCTGCCCCAGCCACTCAAGAGAGGCTCAGAGTGGTGGCC 3759

Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCTGTGACCTCGAGGTGGACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCTGTGACCTCGAGGTGGACCA 3819

Qy 253 CCCTCAAGAGCCTGAGCAGCAGATCGAGAACATCCGGAGCCAGAGGGAGCGCAAGA 312
Db 3820 CCCTCAAGAGCCTGAGCAGCAGATCGAGAACATCCGGAGCCAGAGGGAGCGCAAGA 3879

Qy 313 ACCCCGCCGACCTGCGGTGAACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAT 372
Db 3880 ACCCCGCCGACCTGCGGTGAACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAT 3939

Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGTGATGCCATCAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGTGATGCCATCAAGTCTTCTGCAACATGG 3999

Qy 433 AGACTGTGTGACCTGCTGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACA 492
Db 4000 AGACTGTGTGACCTGCTGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACA 4059

Qy 493 TCAGCAAGAACCCCAAGGAGAGGATGTCTGGTTCGGCGAGAGATGACCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGGAGAGGATGTCTGGTTCGGCGAGAGATGACCGATGGAT 4119

Qy 553 TCAGTTCGATGTGCGGCGAGGCTCCAGACATCACTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 612
Db 4120 TCAGTTCGATGTGCGGCGAGGCTCCAGACATCACTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 3939

Qy 613 ACCCCGCCGACCTGCGGTGAACCTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 432
Db 4180 ACCCCGCCGACCTGCGGTGAACCTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 3939

Qy 673 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 732
Db 4240 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4299

Qy 732 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 732
Db 4299 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4299

Qy 792 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 792
Db 4300 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4359

Qy 852 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 852
Db 4360 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4419

Qy 912 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 912
Db 4420 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4479

Qy 959 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 959
Db 4480 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4526

RESULT 8
AX393330
LOCUS
DEFINITION Sequence 260 from Patent WO0210217.
ACCESSION AX393330

AX393330 6728 bp DNA linear PAT 23-MAR-2002

VERSION * AX393330.1 GI:19701337
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.

REFERENCE 1
AUTHORS St Croix, B., Kinzler, K.W. and Vogelstein, B.
TITLE Endothelial cell expression patterns
JOURNAL Patent: WO 0210217-A 260 07-FEB-2002;
The Johns Hopkins University (US)
FEATURES
source 1. .6728
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 97.8%; Score 942.2; DB 6; Length 6728;
Best Local Similarity 99.7%; Pred. No. 5.1e-138;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 13 GATCTAAGGTCCTCCCTGGGCCCCATTGGGCCCCCTGGTCTCGGGTTCGCACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGGCCCCATTGGGCCCCCTGGTCTCGGGTTCGCACTGGTGATG 3639

Qy 73 CTGGTCTCTGTTGGTCCCTCCGGCCCTCTCGGACCTCTCGGTCCTCCCTGGTCTCCCTCCAGCG 132
Db 3640 CTGGTCTCTGTTGGTCCCTCCGGCCCTCTCGGACCTCTCGGTCCTCCCTGGTCTCCCTCCAGCG 3699

Qy 133 CTGGTCTCTGATTCAGCTTCTCTGCCCCAGCCACTCAAGAGAGGCTCAGAGTGGTGGCC 192
Db 3700 CTGGTCTCTGATTCAGCTTCTCTGCCCCAGCCACTCAAGAGAGGCTCAGAGTGGTGGCC 3759

Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCTGTGACCTCGAGGTGGACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCTGTGACCTCGAGGTGGACCA 3819

Qy 253 CCCTCAAGAGCCTGAGCAGCAGATCGAGAACATCCGGAGCCAGAGGGAGCGCAAGA 312
Db 3820 CCCTCAAGAGCCTGAGCAGCAGATCGAGAACATCCGGAGCCAGAGGGAGCGCAAGA 3879

Qy 313 ACCCCGCCGACCTGCGGTGAACCTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 372
Db 3880 ACCCCGCCGACCTGCGGTGAACCTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 3939

Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGTGATGCCACTCAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGTGATGCCACTCAAGTCTTCTGCAACATGG 3999

Qy 433 AGACTGTGTGACCTGCTGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACA 492
Db 4000 AGACTGTGTGACCTGCTGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACA 4059

Qy 493 TCAGCAAGAACCCCAAGGAGAGGATGTCTGGTTCGGCGAGAGATGACCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGGAGAGGATGTCTGGTTCGGCGAGAGATGACCGATGGAT 4119

Qy 553 TCAGTTCGATGTGCGGCGAGGCTCCAGACATCACTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 612
Db 4120 TCAGTTCGATGTGCGGCGAGGCTCCAGACATCACTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 4179

Qy 613 ACCCCGCCGACCTGCGGTGAACCTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 612
Db 4180 ACCCCGCCGACCTGCGGTGAACCTCAAGATGTGCCACTCTGAGAGAGTGGAGAT 4179

Qy 673 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 732
Db 4240 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4299

Qy 732 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 732
Db 4299 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4299

Qy 792 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 792
Db 4300 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4359

Qy 852 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 852
Db 4360 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4419

Qy 912 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 912
Db 4420 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4479

Qy 959 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 959
Db 4480 TGGCTTACATGGACAGCAGCTGGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCCA 4526

```
Db 4300 ACGAGATCGAGATCCGCGCCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTCTGATG 4359
Qy 793 GCTGCAAGAGTCAACACGAGCCCTGGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGT 852
Db 4360 GCTGCAAGAGTCAACACGAGCCCTGGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGT 4419
Qy 853 CCTCCCGCTGSCCATCATCATGATGTGCCCCCTTGGAGCGTTGGTGCCGACAGCAGGAAT 912
Db 4420 CCTCCCGCTGSCCATCATCATGATGTGCCCCCTTGGAGCGTTGGTGCCGACAGCAGGAAT 4479
Qy 913 TCGGCTTCGAGCTTGGGCGCTCTGCTTCTCTGTAAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGCTTGGGCGCTCTGCTTCTCTGTAAACTCCCTCCATC 4526

RESULT 9
AX411302
LOCUS AX411302 6728 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 3949 from Patent WO0229103.
ACCESSION AX411302
VERSION AX411302.1 GI:21444007
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
TITLE Gene expression profiles in liver cancer
JOURNAL Patent: WO 0229103-A 3949 11-APR-2002;
GENE LOGIC INC (US)
FEATURES
    source      1..6728
                Location/Qualifiers
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
                /note="EMBL/GenBank Accession No. Z74615"

ORIGIN
Query Match 97.8%; Score 942.2; DB 6; Length 6728;
Best Local Similarity 99.7%; Pred. No. 5.1e-138;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 13 GATCTAACGGTCTCCCTGGCCCCATTGGGGCCCCCTGTCTCGCGGTGCGACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGCCCCATTGGGGCCCCCTGTCTCGCGGTGCGACTGGTGATG 3639
Qy 73 CTGCTCTCTGTTGGTCCCGCCCGCCCTCTGTGACCTCTCTGTCCTCCCTGCTCCCGACGG 132
Db 3640 CTGCTCTCTGTTGGTCCCGCCCGCCCTCTGTGACCTCTCTGTCCTCCCTGCTCCCGACGG 3699
Qy 133 CTGGTTTCGACTTCAGCTTCTGTCGCCCGCCAGCCCTCAAGAGAGGCTCAGGATGGTGCC 192
Db 3700 CTGGTTTCGACTTCAGCTTCTGTCGCCCGCCAGCCCTCAAGAGAGGCTCAGGATGGTGCC 3759
Qy 193 GCTACTACTCCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 252
Db 3760 GCTACTACTCCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3819
Qy 253 CCTCTCAAGAGCTGAGCCAGAGATCGAGAACATCTCGGAGCCCGAGAGGAGCGCGAAGA 312
Db 3820 CCTCTCAAGAGCTGAGCCAGAGATCGAGAACATCTCGGAGCCCGAGAGGAGCGCGAAGA 3879
Qy 313 ACCCGCCCGCAGCTGCGGTGACCTCAAGATGTCGCACTCTGACTGGAAGGTGGAGAT 372
Db 3880 ACCCGCCCGCAGCTGCGGTGACCTCAAGATGTCGCACTCTGACTGGAAGGTGGAGAT 3939
Qy 373 ACTGATTTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 432
Db 3940 ACTGATTTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 3999
Qy 433 AGACTGTGTGAGACTGTGCTGTGTACCCCACTCAGGCCAGTGTGGCCCCAGAGAACTGGTACA 492
```

```
Db 4000 AGACTGTGTGAGACTGTGCTGTATCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACA 4059
Qy 493 TCAGCAAGAAACCCCAAGGCAAGAGGCAATCTGTGGTTTCGGCGAGAGCATCAGCCGATGGAT 552
Db 4060 TCAGCAAGAAACCCCAAGGCAAGAGGCAATCTGTGGTTTCGGCGAGAGCATCAGCCGATGGAT 4119
Qy 553 TCAGTTTCGAGTATGGGGGCGAGGCTTCGCAACCTGCGCGATGTGGCCATCAGCTGACCT 612
Db 4120 TCAGTTTCGAGTATGGGGGCGAGGCTTCGCAACCTGCGCGATGTGGCCATCAGCTGACCT 4179
Qy 613 TCCTGGGCTGATGTCCACCGAGGCTTCGCAAGCAATCACTACCACTGCAAGAAACAGCG 672
Db 4180 TCCTGGGCTGATGTCCACCGAGGCTTCGCAAGCAATCACTACCACTGCAAGAAACAGCG 4239
Qy 673 TGGCTTACATGACGACGACTGGCAACCTCAAGAGGCGCTGCTCTCTCAAGGGCTTCCA 732
Db 4240 TGGCTTACATGACGACGACTGGCAACCTCAAGAGGCGCTGCTCTCTCAAGGGCTTCCA 4299
Qy 733 ACGAGATCGAGATCCGGGCGAGGCGCAACGCGCTTCACCTACAGGCTCACTGTCTGATG 792
Db 4300 ACGAGATCGAGATCCGGGCGAGGCGCAACGCGCTTCACCTACAGGCTCACTGTCTGATG 4359
Qy 793 GCTGCAAGAGTCAACCGAGGCTTCGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGT 852
Db 4360 GCTGCAAGAGTCAACCGAGGCTTCGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGT 4419
Qy 853 CTCTCCGCTTCCGCTCATCATCATGATGTGGCCCCCTTGGACGTTGGTGGCCCGACAGCAAGT 912
Db 4420 CTCTCCGCTTCCGCTCATCATCATGATGTGGCCCCCTTGGACGTTGGTGGCCCGACAGCAAGT 4479
Qy 913 TCGGCTTCGAGCTTGGGCGCTCTGCTTCTCTGTAAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGCTTGGGCGCTCTGCTTCTCTGTAAACTCCCTCCATC 4526

RESULT 10
HSPAL1ICO
LOCUS H.sapiens mRNA for prepro-alpha1(I) collagen. PRI 07-MAR-1997
DEFINITION H.sapiens mRNA for prepro-alpha1(I) collagen.
ACCESSION Z74615
VERSION Z74615.1 GI:1418927
KEYWORDS alpha1(I)-collagen.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1537 to 3803)
AUTHORS Bernard,M.P., Chu,M.L., Myers,J.C., Ramirez,F., Eikenberry,E.F. and Prockop,D.J.
TITLE Nucleotide sequences of complementary deoxyribonucleic acids for the pro alpha 1 chain of human type I procollagen. Statistical evaluation of structures that are conserved during evolution
JOURNAL Biochemistry 22 (22), 5213-5223 (1983)
PUBMED 6889127
REFERENCE 2 (bases 1 to 36)
AUTHORS Chu,M.L., de Wet,W., Bernard,M. and Ramirez,F.
TITLE Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter structure, AluI repeats, and polymorphic transcripts
JOURNAL J. Biol. Chem. 260 (4), 2315-2320 (1985)
PUBMED 2857713
REFERENCE 3 (bases 3804 to 4481)
AUTHORS Makela,J.K., Raassina,M., Virta,A. and Vuorio,E.
TITLE Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide domain
JOURNAL Nucleic Acids Res. 16 (1), 349 (1988)
PUBMED 3340531
REFERENCE 4 (bases 37 to 1536)
AUTHORS Tromp,G., Kuivaniemi,H., Stacey,A., Shikata,H., Baldwin,C.T., Jaenisch,R. and Prockop,D.J.
TITLE Structure of a full-length cDNA clone for the prepro alpha 1(I) chain of human type I procollagen
```



```
QY 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGG 3999
QY 433 AGACTGGTGAGACTCGGTGTACCCCACTCAGCCCAAGTGTGGCCAGAGAAGTGGGTACA 492
Db 4000 AGACTGGTGAGACTCGGTGTACCCCACTCAGCCCAAGTGTGGCCAGAGAAGTGGGTACA 4059
QY 493 TCAGCAAGAACCCCAAGGCAAGAGGCGATGCTGTGGTGGCGAGAGCATGACCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGGCAAGAGGCGATGCTGTGGTGGCGAGAGCATGACCGATGGAT 4119
QY 553 TCCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTTCCGATGTGGCCATCCAGCTGACCT 612
Db 4120 TCCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTTCCGATGTGGCCATCCAGCTGACCT 4179
QY 613 TCCTGGCGCTCATGTTCACCCAGAGCTCCCAAGAACATCACCCTACCACTGCAAGAACAGCG 672
Db 4180 TCCTGGCGCTCATGTTCACCCAGAGCTCCCAAGAACATCACCCTACCACTGCAAGAACAGCG 4239
QY 673 TGGCTCATACATGAGCAGCAGACTGGCAACCTCAAGAGGGCCCTGCTCTCTCAAGGGCTCCA 732
Db 4240 TGGCTCATACATGAGCAGCAGACTGGCAACCTCAAGAGGGCCCTGCTCTCTCAAGGGCTCCA 4299
QY 733 ACGAGATCGAGATCCGGCCGAGGGGCAACAGCCGCTTCACTACAGCGTCACTGTCGATG 792
Db 4300 ACGAGATCGAGATCCGGCCGAGGGGCAACAGCCGCTTCACTACAGCGTCACTGTCGATG 4359
QY 793 GCTCAGCAGTFCACACCGAGCCCTGGGGCAGACAGTGTATTGATACAAACACCAACCAAGT 852
Db 4360 GCTCAGCAGTFCACACCGAGCCCTGGGGCAGACAGTGTATTGATACAAACACCAACCAAGT 4419
QY 853 CCTCCCGCTGCCCATCATCGATGTGGCCCGCTTGGAGCTTGGTGGCCCGACAGCAAGT 912
Db 4420 CCTCCCGCTGCCCATCATCGATGTGGCCCGCTTGGAGCTTGGTGGCCCGACAGCAAGT 4479
QY 913 TCGGCTTCGAGTTGGCCCTGCTGCTTCTGTTAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTTGGCCCTGCTGCTTCTGTTAACTCCCTCCATC 4526

RESULT 13
AX776246 3347 bp DNA linear PAT 14-JUL-2003
LOCUS Sequence 16 from Patent WO03048362.
DEFINITION
ACCESSION AX776246
VERSION AX776246.1 GI:32693902
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE
1 Farrar,G.J., Humphries,P., Millington-Ward,S. and Kenna,P.F.
AUTHORS Suppression of polymeric alleles
TITLE Patent: WO 03048362-A 16 12-JUN-2003;
JOURNAL College of the Holy and Undivided Trinity of Queen Elizabeth (IE)
FEATURES
location/Qualifiers
source 1. .3347
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 97.1%; Score 935.4; DB 6; Length 3347;
Best Local Similarity 99.2%; Pred. NO. 6.7e-137;
Matches 939; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 13 GATCTAAGGTCCTCGTGGCCCGCATGGCCCGCTTGGCTTCTCGGGTGGCACTGGTATG 72
Db 2189 GACTCAACGGTCTCCTGGCCCGCATGGCCCGCTTGGCTTCTCGGGTGGCACTGGTATG 2248
```

```
QY 73 CTGGTCTCTGTGGTCCCGCCCGGCTCTCTGGACCTCTCTGGTCCCGCTTCTCTCCAGCG 132
Db 2249 CTGGTCTCTGTGGTCCCGCCCGGCTCTCTGGACCTCTCTGGTCCCGCTTCTCTCCAGCG 2308
QY 133 CTGGTTCGACTTCAGCTTCCTGCCCCAGCACCTCAAGAGAAAGGCTCAGCATGGTGGCC 192
Db 2309 CTGGTTCGACTTCAGCTTCCTGCCCCAGCACCTCAAGAGAAAGGCTCAGCATGGTGGCC 2368
QY 193 GCTACTACCGGGCTGATGATGCCAAATGTGTGTTCTGTGACCTGACCTCGAGGTGACACCA 252
Db 2369 GCTACTACCGGGCTGATGATGCCAAATGTGTGTTCTGTGACCTGACCTCGAGGTGACACCA 2428
QY 253 CCCTCAAGAGCCTTGAGCCAGCAGATCGAGAAACATCCGGAGCCCAAGAGGAAGCCGCAAGA 312
Db 2429 CCCTCAAGAGCCTTGAGCCAGCAGATCGAGAAACATCCGGAGCCCAAGAGGAAGCCGCAAGA 2488
QY 313 ACCCGCCCGCCACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 372
Db 2489 ACCCGCCCGCCACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 2548
QY 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGG 432
Db 2549 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGG 2608
QY 433 AGACTGGTGAGACTCGGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAACTGGGTACA 492
Db 2609 AGACTGGTGAGACTCGGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAACTGGGTACA 2668
QY 493 TCAGCAAGAACCCCAAGGCAAGAGGCGATGCTGTGGTTCGGCGAGAGCATGACCCGATGGAT 552
Db 2669 TCAGCAAGAACCCCAAGGCAAGAGGCGATGCTGTGGTTCGGCGAGAGCATGACCCGATGGAT 2728
QY 553 TCCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612
Db 2729 TCCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 2788
QY 613 TCTGGCGCTGATGTCCAGAGGCTTCCAGAACATCACTTACCACTGCAAGAACAGCG 672
Db 2789 TCTGGCGCTGATGTCCAGAGGCTTCCAGAACATCACTTACCACTGCAAGAACAGCG 2848
QY 673 TGGCTTACATGAGCAGCAGACTGGCAACCTCAAGAGGGCCCTGCTCTCAAGGGCTCCA 732
Db 2849 TGGCTTACATGAGCAGCAGACTGGCAACCTCAAGAGGGCCCTGCTCTCAAGGGCTCCA 2908
QY 733 ACGAGATCGAGATCCGGCCGAGGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATG 792
Db 2909 ACGAGATCGAGATCCGGCCGAGGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATG 2968
QY 793 GCTGCAAGTTCACACCGGAGCCTGGGGCAAGACAGTGTGAATACAAACACCAACCAAGT 852
Db 2969 GCTGCAAGTTCACACCGGAGCCTGGGGCAAGACAGTGTGAATACAAACACCAACCAAGT 3028
QY 853 CCTCCCGCTGCCCATCATCGATGTGGCCCGCTTGGAGCTTGGTGGCCCGACAGCAAGT 912
Db 3029 CCTCCCGCTGCCCATCATCGATGTGGCCCGCTTGGAGCTTGGTGGCCCGACAGCAAGT 3088
QY 913 TCGGCTTCGAGTTGGCCCTGCTGCTTCTGTTAACTCCCTCCATC 959
Db 3089 TCGGCTTCGAGTTGGCCCTGCTGCTTCTGTTAACTCCCTCCATC 3135

RESULT 14
HUMCG1PAL 3347 bp mRNA linear PRI 01-NOV-1994
LOCUS Human proalpha 1 (I) chain of type I procollagen mRNA (partial).
DEFINITION K01228
ACCESSION K01228.1 GI:180391
VERSION collagen; type I collagen.
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
```


Db	4357	GCTGCACGAGTCACACCGGAGCCTGGGGCAAGACAGTGAATTGAATACAAAACCAACAAGA	4416
Qy	853	CCTCCCGCCTTGCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGCCCCCAGACCAAGGAAT	912
Db	4417	CCTCCCGCCTTGCGCATCATCGATGTGGCCCCCTTGGACGTTGGTGCCCCCAGACCAAGGAAT	4476
Qy	913	TCGGCTTCGACGTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC	959
Db	4477	TCGGCTTCGACGTTGGCCCATGTCTGCTTCTGTAAACTCCCTCCATC	4523

Search completed: April 27, 2006, 20:12:54
Job time : 5492.32 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 17:14:00 ; Search time 602.014 Seconds
(without alignments)
10661.049 Million cell updates/sec

Title: US-10-677-877A-1
Perfect score: 963
Sequence: 1 agcttcacgaagatctaac.....gtaaacctccctccatctaga 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseqn_21:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002as:*
7: Geneseqn2002bs:*
8: Geneseqn2003as:*
9: Geneseqn2003bs:*
10: Geneseqn2003cs:*
11: Geneseqn2003ds:*
12: Geneseqn2004as:*
13: Geneseqn2004bs:*
14: Geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	963	100.0	963	14	Aea28568 Human col
2	956.2	99.3	2487	14	Aea28572 Human alk
3	952	98.9	2139	14	Aea28580 Human sol
4	951	98.8	1734	14	Aea28576 Human sol
5	942.2	97.8	5921	8	Acc50109 Breast ca
6	942.2	97.8	5921	11	Adp65048 Human col
7	942.2	97.8	5921	12	Adq19469 Human sof
8	942.2	97.8	5921	12	Adq29586 Human col
9	942.2	97.8	5921	13	Adr99017 Collagen,
10	942.2	97.8	5921	14	Adz26572 Human typ
11	942.2	97.8	5921	14	Adz09709 Human bre
12	942.2	97.8	5921	14	Aea04387 Human cdn
13	942.2	97.8	6728	4	Aaf90491 Human pro
14	942.2	97.8	6728	6	Abi62184 Colon ade
15	942.2	97.8	6728	6	Abi62729 Colon ade
16	942.2	97.8	6728	6	Abi65472 Lung canc
17	942.2	97.8	6728	6	Abi92119 Human Tum
18	942.2	97.8	6728	6	Abn97451 Gene #394
19	942.2	97.8	6728	6	Abv94755 Human pan

20	942.2	97.8	6728	8	Acc50108	Acc50108 Breast ca
21	942.2	97.8	6728	8	Acfl12845	Acfl12845 Human cer
22	942.2	97.8	6728	10	Add14744	Add14744 Human src
23	942.2	97.8	6728	10	Abx72044	Abx72044 DNA encod
24	942.2	97.8	6728	11	Adp65585	Adp65585 Human mrn
25	942.2	97.8	6728	14	Adv70139	Adv70139 Tumor-as
26	940.6	97.7	4213	13	AdS97949	AdS97949 Protein f
27	937.4	97.3	3232	10	AdS87386	AdS87386 Human pan
28	937.4	97.3	4073	10	AdS87389	AdS87389 Human pan
29	937.4	97.3	4744	10	AdJ56243	AdJ56243 Human cdn
30	937.4	97.3	4770	4	AAS22441	AAS22441 Human cdn
31	937.4	97.3	4784	4	AAS22677	AAS22677 Human cdn
32	937.4	97.3	4791	13	AdS98292	AdS98292 Protein f
33	937.4	97.3	4952	8	Acc46588	Acc46588 Human dit
34	937.4	97.3	4952	8	AdS87382	AdS87382 Human pan
35	937.4	97.3	5017	10	AdS87387	AdS87387 Human pan
36	937.4	97.3	5739	10	AdS87383	AdS87383 Human pan
37	937.4	97.3	5739	12	AdF45431	AdF45431 Human vas
38	937.4	97.3	5919	8	ACA64815	ACA64815 Human col
39	937.4	97.3	5920	12	AdQ86616	AdQ86616 Human tum
40	937.4	97.3	5920	13	ACN40246	ACN40246 Tumour-as
41	937.4	97.3	5930	10	AdS87384	AdS87384 Human pan
42	937.4	97.3	6186	10	AdS87388	AdS87388 Human pan
43	937.4	97.3	6739	9	ACH04032	ACH04032 Human cdn
44	937.4	97.3	6753	11	ACN92102	ACN92102 Breast ca
45	937.4	97.3	6753	11	ACN92102	ACN92102 Breast ca

ALIGNMENTS

RESULT 1
AEA28568
ID AEA28568 standard; cdna; 963 BP.

XX AEA28568;
XX
XX
DT 28-JUL-2005 (first entry)
XX
DE Human collagen alpha(I) C-propeptide T0 construct cDNA.
XX
KW protein secretion; fusion protein; protein activation; collagen I; ss;
KW gene.
XX
OS Homo sapiens.
OS Synthetic.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 12..947
FT /tag= a
FT /partial
FT /product= "Human collagen alpha(I) C-propeptide T0
FT construct protein"
FT /note= "No start codon"

WO2005047850-A2.
26-MAY-2005.
04-OCT-2004; 2004WO-US032753.
02-OCT-2003; 2003US-00677877.
(GENH-) GENHUNTER CORP.
Liang P;
WPI; 2005-386406/39.
P-PSDB; AEA28569.

Generating a secreted trimeric fusion protein by introducing into a cell a DNA construct comprising a promoter linked to a template encoding a signal peptide sequence followed by in-frame fusion to polypeptide to be

trimeric fusion protein. The method comprises creating a DNA construct comprising a transcriptional promoter linked to a template encoding a signal peptide sequence followed, via an in-frame fusion, to the polypeptide to be trimerized, which in turn is joined in-frame to a polypeptide capable of self-trimerization which is heterologous to the first polypeptide to be trimerized, and introducing the construct into a eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of only the extracellular ligand-binding domain, may be used to intercept a ligand and thus overcome the overactivation of a receptor during disease treatment. Trimeric receptor decoys theoretically should have a much higher affinity to their ligand than their dimeric counterparts. Such rationally designed soluble trimeric receptor analogs could significantly increase clinical benefits, as well as lower the amount or frequency of drug injections for each patient. The method of the invention may be useful for generating a secreted trimeric fusion protein to be used for more efficient neutralization of the biological activities of their naturally occurring trimeric ligands. The current sequence is that of the human placental alkaline phosphatase (AP)-human collagen alpha(I) C-propeptide T0 fusion DNA (AP-T0) of the invention. The collagen T0 construct includes a partial glycine-repeat triple helical region.

XX SQ Sequence 2487 BP; 525 A; 813 C; 728 G; 421 T; 0 U; 0 Other;

Query Match 99.1%; Score 956.2; DB 14; Length 2487;

Best Local Similarity 99.7%; Pred. No. 1.2e-194;

Matches 958; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	3	GCTTACGTAAAGATCTAAGCGTCTCCCTGGCCCAATGGGCCCCCTCGTCTCGCGTGGC	62
DB	1527	GGTTCGGGAAGATCTAAGCGTCTCCCTGGCCCAATGGGCCCCCTCGTCTCGCGTGGC	1586
QY	63	ACTGGTGAATGCTGCTCTGTTGGTCTCCCGGCCCCCTCGTGGACCTCTGGTCCCTCGT	122
DB	1587	ACTGGTGAATGCTGCTCTGTTGGTCTCCCGGCCCCCTCGTGGACCTCTGGTCCCTCGT	1646
QY	123	CTCTCCAGCGCTGTTTTCGACTTTCAGCTTCTCTCCCGGCCCCCTCGTGGACCTCTGGT	182
DB	1647	CTTCCAGCGCTGTTTTCGACTTTCAGCTTCTCTCCCGGCCCCCTCGTGGACCTCTGGT	1706
QY	183	GATGTTGGCGCTACTACCGGCTGATGATGCCAATGTGTTCTGTCGACCGTGCCTCGAG	242
DB	1707	GATGTTGGCGCTACTACCGGCTGATGATGCCAATGTGTTCTGTCGACCGTGCCTCGAG	1766
QY	243	GTGGACACCACTCTAAGAGCTTGAGCAGCAGATCGAGAACATCCGAGGCCACAGAGGA	302
DB	1767	GTGGACACCACTCTAAGAGCTTGAGCAGCAGATCGAGAACATCCGAGGCCACAGAGGA	1826
QY	303	AGCCGCAAGAACCCCGCCGACCTCCGTCGACCTCAAGATGTCACCTCTGACTGCAAG	362
DB	1827	AGCCGCAAGAACCCCGCCGACCTCCGTCGACCTCAAGATGTCACCTCTGACTGCAAG	1886
QY	363	AGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTC	422
DB	1887	AGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTC	1946
QY	423	TGCAACATGGAGACTGGTGGAGACTCGGTGTATCCCACTCAGCCAGTGTGGCCCAAG	482
DB	1947	TGCAACATGGAGACTGGTGGAGACTCGGTGTATCCCACTCAGCCAGTGTGGCCCAAG	2006
QY	483	AACTGGTACATCAGCAAGAACCCCAAGGACCAAGGACATGCTGTTCGGCGAGAGCATG	542
DB	2007	AACTGGTACATCAGCAAGAACCCCAAGGACCAAGGACATGCTGTTCGGCGAGAGCATG	2066
QY	543	ACCGATGGATTCCAGTTTCAGATTGGCGGCGAGGCTCCGACCCCTGCCGATGTGGCCATC	602
DB	2067	ACCGATGGATTCCAGTTTCAGATTGGCGGCGAGGCTCCGACCCCTGCCGATGTGGCCATC	2126
QY	603	CAGCTGACCTTCTGCGCTGATGTCCAGAGGCTCCGAGAACATCACTACCACTGC	662
DB	2127	CAGCTGACCTTCTGCGCTGATGTCCAGAGGCTCCGAGAACATCACTACCACTGC	2186
QY	663	AAGAACAGGTGCTCTACATGGACCAAGCAGACTGGCAACTCAAGAGGCCCTGCTCTC	722

Db	2187	AAGAACAGCGTGGCTTACATGGACCAAGCAGACTGGCAACCTCAAGAAGGCCCTGCTCTC	2246
QY	723	AAGGGCTCCAAAGAGATCGAGATCCCGCGCGAGGGCAACAGCGCTTCACTACAGGTC	782
Db	2247	AAGGGCTCCAAAGAGATCGAGATCCCGCGCGAGGGCAACAGCGCTTCACTACAGGTC	2306
QY	783	ACTGTGATGGCTGACAGAGTCAACCGGAGCCTGGGGCAAGACAGTGTGAATCAAA	842
Db	2307	ACTGTGATGGCTGACAGAGTCAACCGGAGCCTGGGGCAAGACAGTGTGAATCAAA	2366
QY	843	ACCACCAAGTCTCCCGCCTGCGCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCA	902
Db	2367	ACCACCAAGTCTCCCGCCTGCGCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCA	2426
QY	903	GACCAGGAATGGCTTTCGACGTTGGCCCTGTCTGCTTCTGTAAACTCCCTCATCTAG	962
Db	2427	GACCAGGAATGGCTTTCGACGTTGGCCCTGTCTGCTTCTGTAAACTCCCTCATCTAG	2486
QY	963	A 963	
Db	2487	A 2487	

RESULT 3
AEA28580
ID AEA28580 standard; cDNA; 2139 BP.
XX
AC AEA28580;
XX AC
DT 28-JUL-2005 (first entry)
XX
DE Human soluble CD4-collagen alpha(I) C-propeptide T0 fusion cDNA.
XX
KW protein secretion; fusion protein; protein activation; collagen I; ss;
KW gene; CD4; gene fusion.
OS Homo sapiens.
OS Synthetic.
OS Chimeric.
XX
XX Key Location/Qualifiers
FT CDS 24..2123
FT /tag= a
FT /product= "Human soluble CD4-collagen alpha(I) C-
FT propeptide T0 fusion protein"
XX
PN WO2005047850-A2.
XX
XX
PD 26-MAY-2005.
XX
PF 04-OCT-2004; 2004WO-US032753.
XX
PR 02-OCT-2003; 2003US-00677877.
XX
PA (GENH-) GENHUNTER CORP.
XX
XX Liang P;
XX PI
XX WPI; 2005-386406/39.
DR P-PSDB; AEA28581.
XX
PT Generating a secreted trimeric fusion protein by introducing into a cell
PT a DNA construct comprising a promoter linked to a template encoding a
PT signal peptide sequence followed by in-frame fusion to polypeptide to be
PT trimerized.
XX
PS Claim 16; SEQ ID NO 13; 47pp; English.
XX
CC The invention relates to a novel method for generating a secreted
CC trimeric fusion protein. The method comprises creating a DNA construct
CC comprising a transcriptional promoter linked to a template encoding a
CC signal peptide sequence followed, via an in-frame fusion, to the
CC polypeptide to be trimerized, which in turn is joined in-frame to a

CC polypeptide capable of self-trimerization which is heterologous to the
CC first polypeptide to be trimerized, e.g. introducing the construct into a
CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
CC only the extracellular ligand-binding domain, may be used to intercept a
CC ligand and thus overcome the overactivation of a receptor during disease
CC treatment. Trimeric receptor decoys theoretically should have a much
CC higher affinity to their ligand than their dimeric counterparts. Such
CC rationally designed soluble trimeric receptor analogs could significantly
CC increase clinical benefits, as well as lower the amount or frequency of
CC drug injections for each patient. The method of the invention may be
CC useful for generating a secreted trimeric fusion protein to be used for
CC more efficient neutralization of the biological activities of their
CC naturally occurring trimeric ligands. The current sequence is that of the
CC human soluble CD4-human collagen alpha(I) C-propeptide T0 fusion DNA
CC (sCD4-T0) of the invention. The collagen T0 construct includes a partial
CC glycine-repeat triple helical region.

XX
SQ Sequence 2139 BP; 518 A; 621 C; 590 G; 410 T; 0 U; 0 Other;

Query Match 98.9%; Score 952; DB 14; Length 2139;
Best Local Similarity 100.0%; Pred. No. 9,1e-194;
Matches 952; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 AGATCTAACGGTCTCCTCGCCCATTTGGGCCCTCGTGTCTCGCGTCCGACTGGTGAT 71
DB 1188 AGATCTAACGGTCTCCTCGCCCATTTGGGCCCTCGTGTCTCGCGTCCGACTGGTGAT 1247
QY 72 GCTGGTCTCTGTTGGTCCCCCGGCCCTCTCGACCTCTGTCCTCCCTCGTCTCCACG 131
DB 1248 GCTGGTCTCTGTTGGTCCCCCGGCCCTCTCGACCTCTGTCCTCCCTCGTCTCCACG 1307
QY 132 GCTGGTCTCTGACCTTCAGCTTCTCTGCCCCCAGCCACTCAAGAGAGGCTCAGATGGTGGC 191
DB 1308 GCTGGTCTCTGACCTTCAGCTTCTCTGCCCCCAGCCACTCAAGAGAGGCTCAGATGGTGGC 1367
QY 192 CGCTACTACCGGCTGATGATGCAATGTGTGCTGACCGTGAACCTCGAGGTGGACACC 251
DB 1368 CGCTACTACCGGCTGATGATGCAATGTGTGCTGACCGTGAACCTCGAGGTGGACACC 1427
QY 252 ACCCTCAAGAGCTGAGCCAGCAGATCGAGACATCCGAGCCAGAGGAGCCGACG 311
DB 1428 ACCCTCAAGAGCTGAGCCAGCAGATCGAGACATCCGAGCCAGAGGAGCCGACG 1487
QY 312 AACCCCGCCGACCTCGCTGACCTCAAGATGTGCACTCTGACCTGGAAGAGTGGAGAG 371
DB 1488 AACCCCGCCGACCTCGCTGACCTCAAGATGTGCACTCTGACCTGGAAGAGTGGAGAG 1547
QY 372 TACTGGATTGAACCCCAACCAAGCTGCAACCTGGATGCCATCAAACTCTTCTGCAACATG 431
DB 1548 TACTGGATTGAACCCCAACCAAGCTGCAACCTGGATGCCATCAAACTCTTCTGCAACATG 1607
QY 432 GAGACTGGTGAACCTCGCTGTATCCCACTCAGCCAGTGTGCCAGAGAACTGGTAC 491
DB 1608 GAGACTGGTGAACCTCGCTGTATCCCACTCAGCCAGTGTGCCAGAGAACTGGTAC 1667
QY 492 ATCAGCAAGAACCCCAAGCAGAGAGCATGTCTGGTTCCGGCAGAGCATGACCCATGA 551
DB 1668 ATCAGCAAGAACCCCAAGCAGAGAGCATGTCTGGTTCCGGCAGAGCATGACCCATGA 1727
QY 552 TTCCAGTTTCGAGTATGGGCCCAAGGGCTCCGACCTCGACCTGCCGATGTGCCATCCAGCTGACC 611
DB 1728 TTCCAGTTTCGAGTATGGGCCCAAGGGCTCCGACCTCGACCTGCCGATGTGCCATCCAGCTGACC 1787
QY 612 TTCTCGGCTGTATGTCCACCGAGGCTCCAGAACATCACTACCTGCAAGAACAGCAGC 671
DB 1788 TTCTCGGCTGTATGTCCACCGAGGCTCCAGAACATCACTACCTGCAAGAACAGCAGC 1847
QY 672 GTGGCTTACATGACCAAGCAGAGCTGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCC 731
DB 1848 GTGGCTTACATGACCAAGCAGAGCTGCAACCTCAAGAGGCTCTGCTCTCAAGGGCTCC 1907
QY 732 AACGAGATCGAGATCCGCGCCGAGGCAACAGCCGCTTCACTTACAGCGTCACTGTCTGAT 791

DB 1908 AACGAGATCGAGATCCGCGCCGAGGCAACAGCCGCTTCACTACAGCGTCACTGTCTGAT 1967
QY 792 GGCTGCACGAGTCACACCGGAGCCTGGGGCAAGACAGTGAATTGAATACAAAACCCACCAAG 851
DB 1968 GGCTGCACGAGTCACACCGGAGCCTGGGGCAAGACAGTGAATTGAATACAAAACCCACCAAG 2027
QY 852 TCCTCCCGCTGCCCATCATCGATGTGGCCCCCTTTGGACGTTGGTGGTCCCGCAGACCAAGAA 911
DB 2028 TCCTCCCGCTGCCCATCATCGATGTGGCCCCCTTTGGACGTTGGTGGTCCCGCAGACCAAGAA 2087
QY 912 TTCGGCTTCGAGTGTGGCCCCCTGTCTCTCTCTGTTAAACTCCCTCCATCTAGA 963
DB 2088 TTCGGCTTCGAGTGTGGCCCCCTGTCTCTCTCTGTTAAACTCCCTCCATCTAGA 2139

RESULT 4
AEA28576
ID AEA28576 standard; cDNA; 1734 BP.
XX
AC AEA28576;
XX
DT 28-JUL-2005 (first entry)
XX
DE Human soluble TNF-RII-collagen alpha(I) C-propeptide T0 fusion cDNA.
XX
KW protein secretion; fusion protein; protein activation; collagen I; ss;
KW gene; Tumor necrosis factor; TNF-RII; gene fusion.
XX
OS Homo sapiens.
OS Synthetic.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT CDS 18..1718
FT /*tag= a
FT /product= "Human soluble TNF-RII-collagen alpha(I) C-
FT propeptide T0 fusion protein"
XX
PN WO2005047850-A2.
XX
PD 26-MAY-2005.
XX
PF 04-OCT-2004; 2004WO-US032753.
XX
PR 02-OCT-2003; 2003US-00677877.
XX
PA (GENH-) GENHUNTER CORP.
XX
PI Liang P;
XX
DR WPI: 2005-386406/39.
DR P-PSDB; AEA28577.
XX
PT Generating a secreted trimeric fusion protein by introducing into a cell
PT a DNA construct comprising a promoter linked to a template encoding a
PT signal peptide sequence followed by in-frame fusion to polypeptide to be
PT trimerized.
XX
PS Claim 15; SEQ ID NO 9; 47pp; English.
XX
CC The invention relates to a novel method for generating a secreted
CC trimeric fusion protein. The method comprises creating a DNA construct
CC comprising a transcriptional promoter linked to a template encoding a
CC signal peptide sequence followed, via an in-frame fusion, to the
CC polypeptide to be trimerized, which in turn is joined in-frame to a
CC polypeptide capable of self-trimerization which is heterologous to the
CC first polypeptide to be trimerized, and introducing the construct into a
CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
CC only the extracellular ligand-binding domain, may be used to intercept a
CC ligand and thus overcome the overactivation of a receptor during disease
CC treatment. Trimeric receptor decoys theoretically should have a much
CC higher affinity to their ligand than their dimeric counterparts. Such
CC rationally designed soluble trimeric receptor analogs could significantly

CC increase clinical benefits, as well as lower the amount or frequency of
CC drug injections for each patient. The method of the invention may be
CC useful for generating a secreted trimeric fusion protein to be used for
CC more efficient neutralization of the biological activities of their
CC naturally occurring trimeric ligands. The current sequence is that of the
CC human soluble TNF-R1I-human collagen alpha(I) C-propeptide T0 fusion DNA
CC (sTNF-R1I-T0) of the invention. The collagen T0 construct includes a
CC partial glycine-repeat triple helical region.

XX
SQ Sequence 1734 BP; 358 A; 598 C; 477 G; 301 T; 0 U; 0 Other;
Query Match 98.8%; Score 951; DB 14; Length 1734;
Best Local Similarity 100.0%; Pred. No. 1.4e-193; Indels 0; Gaps 0;
Matches 951; Conservative 0; Mismatches 0;

QY 13 GATCTAACGGTCTCTCCCTGGCCCCATTGGCCCCCTGTCTCGGGTCGCACACTGGTGATG 72
DB 784 GATCTAACGGTCTCTCCCTGGCCCCATTGGCCCCCTGTCTCGGGTCGCACACTGGTGATG 843
QY 73 CTGGTCTCTGTGGTCCCGCCCGCCCTCTGGAACCTCTGTCTCCCTCGGTCTCTCCACAGCG 132
DB 844 CTGGTCTCTGTGGTCCCGCCCGCCCTCTGGAACCTCTGTCTCCCTCGGTCTCTCCACAGCG 903
QY 133 CTGGTCTCTGTGGTCCCGCCCGCCCTCTGGAACCTCTGTCTCCCTCGGTCTCTCCACAGCG 192
DB 904 CTGGTCTCTGTGGTCCCGCCCGCCCTCTGGAACCTCTGTCTCCCTCGGTCTCTCCACAGCG 963
QY 193 GCTACTACCGGGCTGATGATGCAATGTGGTTCTGTGACCGTGTGACCTCGAGGTGACACCA 252
DB 964 GCTACTACCGGGCTGATGATGCAATGTGGTTCTGTGACCGTGTGACCTCGAGGTGACACCA 1023
QY 253 CCTCAAGAGCTGAGCCAGCAGATCGAAGCAATCTCGGAGCCCAAGAGGAGCGCAAGA 312
DB 1024 CCTCAAGAGCTGAGCCAGCAGATCGAAGCAATCTCGGAGCCCAAGAGGAGCGCAAGA 1083
QY 313 ACCCGCCCGCAGCTGCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAT 372
DB 1084 ACCCGCCCGCAGCTGCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAT 1143
QY 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTCGATGCGATCCAAAGTCTTCTGCAACATGG 432
DB 1144 ACTGGATTGACCCCAACCAAGGCTGCAACCTCGATGCGATCCAAAGTCTTCTGCAACATGG 1203
QY 433 AGACTGTGTAGACTGTGGTGTATACCCACTCAGCCAGTGTGGCCCAAGAACTGTGTACA 492
DB 1204 AGACTGTGTAGACTGTGGTGTATACCCACTCAGCCAGTGTGGCCCAAGAACTGTGTACA 1263
QY 493 TCAGCAGAACCCCAAGGACAGAGGCACTCTGTGTTGGCGGAGAGCATCACCGATGGAT 552
DB 1264 TCAGCAGAACCCCAAGGACAGAGGCACTCTGTGTTGGCGGAGAGCATCACCGATGGAT 1323
QY 553 TCAGATTTCGAGTATGGCGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612
DB 1324 TCAGATTTCGAGTATGGCGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 1383
QY 613 TCTGCGCCCTGATGTCCACCGAGCCCTCCAGAACTACCTACCTACCTGCAAGAACAGCG 672
DB 1384 TCTGCGCCCTGATGTCCACCGAGCCCTCCAGAACTACCTACCTACCTGCAAGAACAGCG 1443
QY 673 TGGCTCATGTGACACAGCAGCTGGGCACTCTCAAGAGGCGCTGTCTCTCAAGGGCTCCA 732
DB 1444 TGGCTCATGTGACACAGCAGCTGGGCACTCTCAAGAGGCGCTGTCTCTCAAGGGCTCCA 1503
QY 733 ACAGATTTCGAGTATGGCGGCGAGGGCAACAGCGCTTCACTACAGGCTCACTGTGATG 792
DB 1504 ACAGATTTCGAGTATGGCGGCGAGGGCAACAGCGCTTCACTACAGGCTCACTGTGATG 1563
QY 793 GCTGCACAGTTCACACCGAGCCCTGGGGCAAGACAGTGTGAATACAAACACCAAGT 852
DB 1564 GCTGCACAGTTCACACCGAGCCCTGGGGCAAGACAGTGTGAATACAAACACCAAGT 1623
QY 853 CTTCGCCCTGCCCATCATGTGTGGCCCTTGGAGCTGTGGTGGCCCGCAGACAGGAAT 912

DB 1624 CTTCGCCCTGCCCATCATCATGATGTGGCCCCCTTGGAGTTGGTCCCGCAGACAGGAAT 1683
QY 913 TCGGCTTCGACGTTGGCCCCCTGTCTGTCTCTGTAACCTCCTCATCTAGA 963
DB 1684 TCGGCTTCGACGTTGGCCCCCTGTCTGTCTCTGTAACCTCCTCATCTAGA 1734

RESULT 5
ACC50109
ID ACC50109 standard; cDNA; 5921 BP.
XX AC ACC50109;
XX AC ACC50109;
XX 12-JUN-2003 (first entry)
XX Breast cancer associated cDNA sequence SEQ ID NO:66.
XX Human; breast cancer; cytostatic; gene therapy; gene; ss.
XX Homo sapiens.
XX WO2003004989-A2.
XX 16-JAN-2003.
XX 21-JUN-2002; 2002WO-US019669.
XX 21-JUN-2001; 2001US-0299887P.
XX 27-JUN-2001; 2001US-0301572P.
XX 18-JUL-2001; 2001US-0306501P.
XX 25-SEP-2001; 2001US-0325002P.
XX 05-MAR-2002; 2002US-0362585P.
XX 14-MAY-2002; 2002US-0380391P.
XX (MILL-) MILLENIUM PHARM INC.
XX Lillie J, Ganavarapu M, Glatt K, Hoersh S, Kamatkar S;
XX Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
XX Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
XX WPI; 2003-210381/20.
XX Breast cancer diagnosis or treatment by comparing the level of expression
XX of a marker in a patient sample with that in the control non-breast
XX cancer sample.
XX Claim 1; SEQ ID NO 66; 128pp; English.
XX The present invention describes a method for assessing whether a patient
XX is afflicted with breast cancer. The method comprises comparing the level
XX of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
XX ABR47386 to ABR47632) in a patient sample and the normal level of
XX expression of the marker in a control non-breast cancer sample, where a
XX significant increase in the level of expression of the marker in the
XX patient sample and the normal level is an indication that the patient is
XX afflicted with breast cancer. The breast cancer associated sequences from
XX the present invention have cytostatic activities and can be used in gene
XX therapy. The method is useful for diagnosing and treating breast cancer.
XX N.B. The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
Query Match 97.8%; Score 942.2; DB 8; Length 5921;
Best Local Similarity 99.7%; Pred. No. 1.4e-191; Indels 0; Gaps 0;
Matches 944; Conservative 0; Mismatches 3;

QY 13 GATCTAACGGTCTCTCCCTGGCCCCCATTTGGCCCCCTGTCTCGGGTCGCACACTGGTGATG 72
DB 3580 GATCTAACGGTCTCTCCCTGGCCCCCATTTGGCCCCCTGTCTCGGGTCGCACACTGGTGATG 3639
QY 73 CTGGTCTCTGTGGTCCCGCCCGCCCTCTGGAACCTCTCTGTGGTCCCGCAGACAGGAAT 132

```
Db 3640 CTGGTCTGTGGTCCCGCCCTCTCGACCTCTGTGTCCTCCCTGCTCTCCACGG 3699
Qy 133 CTGGTTTCAGACTTCAGTTCCTGCGCCCGACCTCAAGAGAGGCTCAGATGGTGGCC 192
Db 3700 CTGGTTTCAGACTTCAGTTCCTGCGCCCGACCTCAAGAGAGGCTCAGATGGTGGCC 3759
Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGTGACCTCGAGGTGGAACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGTGACCTCGAGGTGGAACCA 3819
Qy 253 CCTCAAGAGCTTGAGCCAGAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGA 312
Db 3820 CCTCAAGAGCTTGAGCCAGAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGA 3879
Qy 313 ACCCGCCGCGACCTGCGGTGACCTCAAGATGTGCCACTGTGACTGGAAGAGTGGAGAT 372
Db 3880 ACCCGCCGCGACCTGCGGTGACCTCAAGATGTGCCACTGTGACTGGAAGAGTGGAGAT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGTGATGCCATCAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGTGATGCCATCAAGTCTTCTGCAACATGG 3999
Qy 433 AGACTGGTGAAGCTCGGTGTACCCCACTCAGCCCGAGTGTGGCCCAAGAGAACTGTGTA 492
Db 4000 AGACTGGTGAAGCTCGGTGTACCCCACTCAGCCCGAGTGTGGCCCAAGAGAACTGTGTA 4059
Qy 493 TCAGCAAGAACCCCAAGAGCAAGAGCATGTCTGGTTCGCGAGAGCATGCCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGAGCAAGAGCATGTCTGGTTCGCGAGAGCATGCCGATGGAT 4119
Qy 553 TCCAGTTTCAGATGTGGCGGCAGGGCTCCGACCTCGCATGTGGCCATCCAGCTGACCT 612
Db 4120 TCCAGTTTCAGATGTGGCGGCAGGGCTCCGACCTCGCATGTGGCCATCCAGCTGACCT 4179
Qy 613 TCTCGCCTGTGATGCCACGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCG 672
Db 4180 TCTCGCCTGTGATGCCACGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCG 4239
Qy 673 TGGCCTACATGACAGCAGCATGGCAACTCAAGAGGCCCTGTCTCTCAGGGGTCCA 732
Db 4240 TGGCCTACATGACAGCAGCATGGCAACTCAAGAGGCCCTGTCTCTCAGGGGTCCA 4299
Qy 733 ACAGATTCAGATTCGCGCGGAGGCAACAGCCGCTTCACTACAGCGTCACCTGTGATG 792
Db 4300 ACAGATTCAGATTCGCGCGGAGGCAACAGCCGCTTCACTACAGCGTCACCTGTGATG 4359
Qy 793 GCTGCAAGATGTCACCGAGGCTTGGGGCAAGACAGTGAATTGAATCAAAACCAACAGT 852
Db 4360 GCTGCAAGATGTCACCGAGGCTTGGGGCAAGACAGTGAATTGAATCAAAACCAACAGT 4419
Qy 853 CTCCCGCCTGCCCATCATGATGTGGCCCGCTTGGACGTTGGTGGCCCGAGACCGAAT 912
Db 4420 CTCCCGCCTGCCCATCATGATGTGGCCCGCTTGGACGTTGGTGGCCCGAGACCGAAT 4479
Qy 913 TCGGCTTCGAGGTTGGCCCTGTCTGTCTCTGTAAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGGTTGGCCCTGTCTGTCTCTGTAAACTCCCTCCATC 4526
```

RESULT 6

```
ADP65048
ID ADP65048 standard; DNA; 5921 BP.
XX
AC ADP65048;
XX
XX 12-AUG-2004 (first entry)
XX
XX Human collagen, type I, alpha 1 (COL1A1) DNA sequence.
XX
XX autoimmune disease; arthritis; gene expression analysis;
KW rheumatoid arthritis; collagen-induced; immunosuppressive; antiarthritis;
KW antiarthritis; osteopathic; antitumor; antitumor; antitumor; dermatological;
```

```
KW immunomodulatory; lupus; ankylosing spondylitis; Fibrositis;
KW fibromyalgia; osteoarthritis; gout; juvenile rheumatoid arthritis;
KW immune; ds; human.
OS Homo sapiens.
PN WO2003072827-A1.
XX 04-SEP-2003.
XX 31-OCT-2002; 2002WO-US035433.
XX 31-OCT-2001; 2001US-0336220P.
XX (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX Hirsch R, Thorton SL;
XX WPI; 2003-712740/67.
XX GENBANK; NM_000088.
XX
XX Diagnosing and analyzing autoimmune disease using gene expression
XX profiles and microarray technology, useful for diagnosing and treating
XX rheumatoid arthritis, lupus, fibrositis, osteoarthritis, fibromyalgia and
XX gout.
XX Disclosure; Page; 56pp; English.
XX
XX The invention relates to a novel method for diagnosing and analyzing
XX autoimmune disease or arthritides. The method comprises obtaining a
XX patient sample containing mRNA, analysing gene expression using the mRNA
XX that results in a gene expression signature of the mRNA, and using that
XX gene expression signature to diagnose or analyse the autoimmune disease
XX or arthritides in the patient, where gene expression of at least 60% of
XX the genes correlates with that of the gene signature. The invention
XX further comprises: a treatment of rheumatoid arthritis; identification of
XX genes for targeting in the treatment of rheumatoid arthritis in a mammal
XX other than a mouse; diagnosis of rheumatoid arthritis in a mammal; an
XX array or gene chip, specific for rheumatoid arthritis; diagnosis or
XX analyses of autoimmune disease or rheumatoid arthritis; screening the
XX efficacy of a candidate drug in vitro for the treatment of collagen-
XX induced arthritis; and reducing the symptoms associated with collagen-
XX induced arthritis. The compositions of the invention have the following
XX activities: immunosuppressive, antirheumatic, antiarthritic, osteopathic,
XX antitumor, antinflammatory, dermatological, and immunomodulatory. The
XX methods and compositions of the present invention are useful for
XX diagnosing and treating autoimmune disease or arthritides, such as
XX rheumatoid arthritis, lupus, ankylosing spondylitis, fibrositis,
XX fibromyalgia, osteoarthritis, gout, juvenile rheumatoid arthritis, and an
XX immune disease caused by an infectious agent. This polynucleotide
XX represents a DNA sequence relating to the genes used in the analysis and
XX treatment of autoimmune diseases or arthritides. Note: This sequence is
XX not shown in the specification. It has been supplied in an electronic
XX format from WIPO.
XX
XX Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
```

```
Query Match 97.8%; Score 942.2; DB 11; Length 5921;
Best Local Similarity 99.7%; Pred. No. 1.4e-191;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 13 GATCTAACGGTCTCCCTGCGCCCATTTGGGCCCCCTGGTCTCTCGCGTCCGACTGGTATG 72
Db 3580 GACTCAACGGTCTCCCTGCGCCCATTTGGGCCCCCTGGTCTCTCGCGTCCGACTGGTATG 3639
Qy 73 CTGGTCTCTGTGGTTCCTCCCGGCTTCCTGACCTCTCTGGTCCCTCTGCTCCACGG 132
Db 3640 CTGGTCTCTGTGGTTCCTCCCGGCTTCCTGACCTCTCTGGTCCCTCTGCTCCACGG 3699
Qy 133 CTGGTTCCTGACTTCAGCTTCCTGCGCCCGACCTCAAGAGAGGCTCAGATGGTGGCC 192
Db 3700 CTGGTTCCTGACTTCAGCTTCCTGCGCCCGACCTCAAGAGAGGCTCAGATGGTGGCC 3759
```

```
QY 193 GCTACTACCGGCTGATGATGCCAATGTTGTTCTGACCGTGACCTCGAGTGGACACCA 252
Db |||||
QY 3760 GCTACTACCGGCTGATGATGCCAATGTTGTTCTGACCGTGACCTCGAGTGGACACCA 3819
Db |||||
QY 253 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 312
Db |||||
QY 3820 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 3879
Db |||||
QY 313 ACCCGCCCGGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 372
Db |||||
QY 3880 ACCCGCCCGGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 3939
Db |||||
QY 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTTCGAACATGG 432
Db |||||
QY 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTTCGAACATGG 3999
Db |||||
QY 433 AGACTGTGAGACCTGCGGTGTACCCCACTCAGCCCACTGAGCCCGAGAACCTGGTACA 492
Db |||||
QY 4000 AGACTGTGAGACCTGCGGTGTACCCCACTCAGCCCACTGAGCCCGAGAACCTGGTACA 4059
Db |||||
QY 493 TCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGTTGGCGAGAGCATGACCGATGGAT 552
Db |||||
QY 4060 TCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGTTGGCGAGAGCATGACCGATGGAT 4119
Db |||||
QY 553 TCAGTTTCGAGTATGGCGGCGAGGCTCCGACCTGCGGATGTGGCCATCCAGTGAACCT 612
Db |||||
QY 4120 TCAGTTTCGAGTATGGCGGCGAGGCTCCGACCTGCGGATGTGGCCATCCAGTGAACCT 4179
Db |||||
QY 613 TCCTGCGCCTGATGTCCACCGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCG 672
Db |||||
QY 4180 TCCTGCGCCTGATGTCCACCGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCG 4239
Db |||||
QY 673 TGGCCTACATGGACCAAGCAGATGGCAACTCAAGAAAGGCTCTCTCAAGGGCTTCCA 732
Db |||||
QY 4240 TGGCCTACATGGACCAAGCAGATGGCAACTCAAGAAAGGCTCTCTCAAGGGCTTCCA 4299
Db |||||
QY 733 ACAGATCGAGATCCGCGCGAGGGCAAGCGCTTCACTACAGGCTCACTGTGATG 792
Db |||||
QY 4300 ACAGATCGAGATCCGCGCGAGGGCAAGCGCTTCACTACAGGCTCACTGTGATG 4359
Db |||||
QY 793 GCTGACAGGTCAACCGGAGCCTGGGGCAAGACAGTGTGATTAATAACAAACCAACCAAGT 852
Db |||||
QY 4360 GCTGACAGGTCAACCGGAGCCTGGGGCAAGACAGTGTGATTAATAACAAACCAACCAAGT 4419
Db |||||
QY 853 CTTCCGCTGCCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCAGACAGGAAT 912
Db |||||
QY 4420 CTTCCGCTGCCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCAGACAGGAAT 4479
Db |||||
QY 913 TCGGCTTCGAGCTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC 959
Db |||||
QY 4480 TCGGCTTCGAGCTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526
Db |||||

RESULT 7
ADQ19469
ID ADQ19469 standard; DNA; 5921 BP.
XX
AC ADQ19469;
XX
DT 26-AUG-2004 (first entry)
XX
DE Human soft tissue sarcoma-upregulated DNA - SEQ ID 2288.
XX
KW soft tissue sarcoma; cytotstatic; gene therapy; vaccine; screening; human;
XX
OS ds.
XX
XX Homo sapiens.
XX
XX WO2004048938-A2.
XX
XX 10-JUN-2004.
XX
XX 26-NOV-2003; 2003WO-US038193.
XX
PF
```

```
XX 26-NOV-2002; 2002US-0429739P.
XX (PROT-) PROTEIN DESIGN LABS INC.
XX Aziz N, Ginsburg WM, Zlotnik A;
XX WPI; 2004-441208/41.
XX
XX Early detection of soft tissue sarcoma comprises determining expression
XX of a gene in a first soft tissue sample and a normal soft tissue sample
XX and comparing the gene expression, also useful in treating soft tissue
XX sarcoma.
XX
XX Example 2; SEQ ID NO 2288; 210pp; English.
XX
XX The invention relates to a novel method for detecting soft tissue sarcoma
XX which comprises obtaining a first soft tissue sample from an individual
XX and a normal soft tissue sample from the same or different individual,
XX determining the expression of a gene in both samples and comparing the
XX expression of the gene in both soft tissue samples, where a higher level
XX of protein expression in the first soft tissue sample indicates the
XX presence of soft tissue sarcoma. The method of the invention has
XX cytotstatic applications and may be useful for detecting soft tissue
XX sarcoma, possibly via gene therapy or vaccine production. The nucleic
XX acid sequences may be useful in diagnostic and screening applications.
XX The current sequence is that of a human soft tissue sarcoma-upregulated
XX DNA of the invention. The current sequence is not shown within the
XX specification per se but was submitted in CD format by the inventor.
XX
XX Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
XX
XX Query Match 97.8%; Score 942.2; DB 12; Length 5921;
XX Best Local Similarity 99.7%; Pred. No. 1.4e-191;
XX Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
QY 13 GATCTAACGGTCTCCCTGGCCCCCATTTGGGCCCCCTGGTCTCTCGGGTGCACACTGGTATG 72
Db |||||
QY 3580 GACTCAACGGTCTCCCTGGCCCCCATTTGGGCCCCCTGGTCTCTCGGGTGCACACTGGTATG 3639
Db |||||
QY 73 CTGGTCTCTGTTGTTCCCGGCCCCCTCTCTGAGCTCTCTGGTCCCTCTGGTCTCTCCAGCG 132
Db |||||
QY 3640 CTGGTCTCTGTTGTTCCCGGCCCCCTCTCTGAGCTCTCTGGTCCCTCTGGTCTCTCCAGCG 3699
Db |||||
QY 133 CTGGTCTCTGAGTTCAGCTTCTCTCCCGGCCCCCTCTCAAGAGAAAGGCTCAAGTGGTGGCC 192
Db |||||
QY 3700 CTGGTCTCTGAGTTCAGCTTCTCTCCCGGCCCCCTCTCAAGAGAAAGGCTCAAGTGGTGGCC 3759
Db |||||
QY 193 GCTACTACCGGCTGATGATGCCAATGTGTTGTTGACCGTGGACCTCGAGTGGACACCA 252
Db |||||
QY 3760 GCTACTACCGGCTGATGATGCCAATGTGTTGTTGACCGTGGACCTCGAGTGGACACCA 3819
Db |||||
QY 253 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 312
Db |||||
QY 3820 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 3879
Db |||||
QY 313 ACCCGCCCGGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 372
Db |||||
QY 3880 ACCCGCCCGGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 3939
Db |||||
QY 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTTCGAACATGG 432
Db |||||
QY 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTTCGAACATGG 3999
Db |||||
QY 433 AGACTGTGAGACCTGCGGTGTACCCCACTCAGCCCACTGAGCCCGAGAACCTGGTACA 492
Db |||||
QY 4000 AGACTGTGAGACCTGCGGTGTACCCCACTCAGCCCACTGAGCCCGAGAACCTGGTACA 4059
Db |||||
QY 493 TCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGTTGGCGAGAGCATGACCGATGGAT 552
Db |||||
QY 4060 TCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGTTGGCGAGAGCATGACCGATGGAT 4119
Db |||||
QY 553 TCAGTTTCGAGTATGGCGGCGAGGCTCCGACCTGCGGATGTGGCCATCCAGTGAACCT 612
Db |||||
```


XX 07-APR-2005.
XX 24-SEP-2004; 2004WO-US031524.
XX 25-SEP-2003; 2003US-0506221P.
XX 08-OCT-2003; 2003US-0509594P.
XX (DAND) DANA FARBER CANCER INST INC.
XX Ritz J, Wu CJ;
XX WPI; 2005-273394/28.
XX P-PSDB; ADZ26573.
XX Detecting lineage-specific cells in a biological sample, useful for
XX determining the clinical outcome of a progenitor cell transfer in a
XX subject, comprises identifying lineage-specific mRNA in the sample.
XX Disclosure; SEQ ID NO 198; 393pp; English.
XX The invention relates to a method of detecting lineage-specific cells in
XX a biological sample which comprises identifying lineage-specific mRNA in
XX the sample. The methods are useful for determining the clinical outcome
XX of a progenitor cell transfer in a subject, and for identifying or
XX quantifying lineage-specific cells. The present sequence represents a
XX human cDNA that encodes a protein used to identify lineage-specific
XX cells.
XX
XX Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
XX
Query Match 97.8%; Score 942.2; DB 14; Length 5921;
Best Local Similarity 99.7%; Pred. No. 1.4e-191;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 13 GATCTAACGGTCTCCCTGGCCCCCATTTGGGCCCCCTTGGTCTCCGCGTGGCACTGGTGATG 72
DB 3580 GACTCAACGGTCTCCCTGGGCCCCCATTTGGGCCCCCTTGGTCTCCGCGTGGCACTGGTGATG 3639
QY 73 CTGGTCTGTGTGTCCTCCGCGGCTCTGGACCTCTGGTCTCCGCGTGGCACTGGTGATG 132
DB 3640 CTGGTCTGTGTGTCCTCCGCGGCTCTGGACCTCTGGTCTCCGCGTGGCACTGGTGATG 3699
QY 133 CTGGTTTCAGCTTCAGCTTCTCCGCGGCTCTGGACCTCTGGTCTCCGCGTGGCACTGGTGATG 192
DB 3700 CTGGTTTCAGCTTCAGCTTCTCCGCGGCTCTGGACCTCTGGTCTCCGCGTGGCACTGGTGATG 3759
QY 193 GCTACTACCGGGCTGATGATGCAATGTGTGTTCTGACCGTGACCTCGAGGTGGACACCA 252
DB 3760 GCTACTACCGGGCTGATGATGCAATGTGTGTTCTGACCGTGACCTCGAGGTGGACACCA 3819
QY 253 CCTCAAGAGCTGAGCGCAGCAGATCGAGAACATCCGAGGCCAGAGGGAGCGGCAAGA 312
DB 3820 CCTCAAGAGCTGAGCGCAGCAGATCGAGAACATCCGAGGCCAGAGGGAGCGGCAAGA 3879
QY 313 ACCCGCGCCGACCTGCGGCTCAAGATGTCCTGCTGACCTGCAAGAGTGGAGAGT 372
DB 3880 ACCCGCGCCGACCTGCGGCTCAAGATGTCCTGCTGACCTGCAAGAGTGGAGAGT 3939
QY 373 ACTGGATTGACCCCAACCAAGGCTGCACTGGATGCCATCAAGTCTTCTGCAACATGG 432
DB 3940 ACTGGATTGACCCCAACCAAGGCTGCACTGGATGCCATCAAGTCTTCTGCAACATGG 3999
QY 433 AGACTGGTGAAGCTGGGTGATCCCACTCAGGCCAGTGTGGCCAGAGAACTGGTACA 492
DB 4000 AGACTGGTGAAGCTGGGTGATCCCACTCAGGCCAGTGTGGCCAGAGAACTGGTACA 4059
QY 493 TCAGCAAGAACCCCAAGGACCAAGAGGATGCTGTGGTTCGGCGGAGAGCATGACCGATGGAT 552
DB 4060 TCAGCAAGAACCCCAAGGACCAAGAGGATGCTGTGGTTCGGCGGAGAGCATGACCGATGGAT 4119
QY 553 TCAGTTCAGATGATGCGGCGCAGGGCTCCGACCTCGCCGATGTGGCCATCCAGTCACT 612
DB 553 TCAGTTCAGATGATGCGGCGCAGGGCTCCGACCTCGCCGATGTGGCCATCCAGTCACT

DB 4120 TCCAGTTCCAGTATGCGCGCCAGGGCTCCGACCCCTGCCGATGTGGCCATCCAGCTGACCT 4179
QY 613 TCCTGCGCTGTGTCTCACCGGAGGCTTCCAGAGAAATCACCTACCTACCTGCAAGAACAGCG 672
DB 4180 TCCTGCGCTGTGTCTCACCGGAGGCTTCCAGAGAAATCACCTACCTACCTGCAAGAACAGCG 4239
QY 673 TGGCCTTACATGAGCAGCAGAGCTGGCAACTCAAGAAAGGCTTCTGCTTCAAGGGTTCGA 732
DB 4240 TGGCCTTACATGAGCAGCAGAGCTGGCAACTCAAGAAAGGCTTCTGCTTCAAGGGTTCGA 4299
QY 733 ACGAGATCAGATCCGCGCGGAGGCAACAGCGCTTCACTACAGGTCACCTGTCGATG 792
DB 4300 ACGAGATCAGATCCGCGCGGAGGCAACAGCGCTTCACTACAGGTCACCTGTCGATG 4359
QY 793 GCTGCAAGAGTCAACCGGAGGCTGGGGCAAGACAGTGAATGAATACAAAACCAACCAAGT 852
DB 4360 GCTGCAAGAGTCAACCGGAGGCTGGGGCAAGACAGTGAATGAATACAAAACCAACCAAGT 4419
QY 853 CTCCCGCTGCCCCATCATCGATGTGGGCCCCCTTGGACGTTGGTGGCCCCCAGACAGGAAT 912
DB 4420 CTCCCGCTGCCCCATCATCGATGTGGGCCCCCTTGGACGTTGGTGGCCCCCAGACAGGAAT 4479
QY 913 TGGCTTTCAGAGTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC 959
DB 4480 TGGCTTTCAGAGTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526
RESULT 11
ADZ09709
ID ADZ09709 standard; DNA; 5921 BP.
XX AC ADZ09709;
XX AC ADZ09709;
DT 30-JUN-2005 (first entry)
XX Human breast cancer marker COL1A1 DNA.
DE neoplasm; tumor marker; cytostatic; breast tumor; gene therapy; ds; gene;
XX COL1A1.
XX Homo sapiens.
OS EP1522594-A2.
PN 13-APR-2005.
XX 30-JUN-2004; 2004EP-00015374.
PF 06-OCT-2003; 2003EP-00022587.
XX (FARB) BAYER HEALTHCARE AG.
XX Munnes M, Bojar H;
PI WPI; 2005-265481/28.
DR P-PSDB; ADZ09874.
DR REFSEQ; NM_000088.
XX Characterizing the state of a neoplastic disease in a subject by
XX comparing the pattern of determined expression levels of marker genes in
XX a biological sample with reference patterns of expression levels.
XX Claim 1; SEQ ID NO 153; 542pp; English.
XX The invention relates to a novel method for characterizing the state of a
XX neoplastic disease in a subject comprising determining the pattern of
XX expression levels of at least 6, 8, 10, 15, 20, 30, 40 or 67 marker genes
XX in a biological sample, comparing the pattern of expression levels with
XX reference patterns of expression levels and characterizing the state of
XX the neoplastic disease in the subject from the outcome of the comparison.
XX The method of the invention demonstrates cytosolic applications and may
XX be useful in characterizing the state of a neoplastic disease in a
XX subject, preferably breast cancer, in gene therapy and for screening for

CC compounds having a curative effect on a neoplastic disease. The current
CC sequence is that of the human breast cancer marker COL1A1 DNA of the
CC invention which is differentially expressed in neoplastic tissue.
XX
SQ Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
Query Match 97.8%; Score 942.2; DB 14; Length 5921;
Best Local Similarity 99.7%; Pred. No. 1.4e-191;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 13 GATCAACGGTCTCCCTGGCCCAATGGGCCCCCTGGTCTCTCGGGTCGCACCTGGTGATG 72
Db |||||
Qy 3580 GACTCAACGGTCTCCCTGGCCCAATGGGCCCCCTGGTCTCTCGGGTCGCACCTGGTGATG 3639
Db |||||
Qy 73 CTGCTCTCTGTGGTCCCGCCGCTCTCTGACCTCTCTGCTGCTCCCTGCTCCCAAGC 132
Db |||||
Qy 3640 CTGCTCTCTGTGGTCCCGCCGCTCTCTGACCTCTCTGCTGCTCCCTGCTCCCAAGC 3699
Db |||||
Qy 133 CTGCTCTCTGACCTCTCTGCTCCCGCCGCTCTCTGACCTCTCTGCTGCTCCCTGCTCCCAAGC 192
Db |||||
Qy 3700 CTGCTCTCTGACCTCTCTGCTCCCGCCGCTCTCTGACCTCTCTGCTGCTCCCTGCTCCCAAGC 3759
Db |||||
Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCCTGACCTCGAGTGGACACCA 252
Db |||||
Qy 3760 GCTACTACCGGGCTGATGATGCCAATGTGGTTCCTGACCTCGAGTGGACACCA 3819
Db |||||
Qy 253 CCCTCAAGAGCTGAGCCAGAGATCGAGAACATCCGGAGCCAGAGGAGCGCAAGA 312
Db |||||
Qy 3820 CCCTCAAGAGCTGAGCCAGAGATCGAGAACATCCGGAGCCAGAGGAGCGCAAGA 3879
Db |||||
Qy 313 ACCCGCCCGCAGCTGCGTCACTCAAGATGTGCTCACTCTGCTGGAAGGTGGAGT 372
Db |||||
Qy 3880 ACCCGCCCGCAGCTGCGTCACTCAAGATGTGCTCACTCTGCTGGAAGGTGGAGT 3939
Db |||||
Qy 373 ACTGATTTGACCCCAACCAAGGCTGCAACCTGGAATGCTCAAGTCTTCTGCAACATGG 432
Db |||||
Qy 3940 ACTGATTTGACCCCAACCAAGGCTGCAACCTGGAATGCTCAAGTCTTCTGCAACATGG 3999
Db |||||
Qy 433 AGACTGTGAGACTGCGTGTGACCCCACTACGAGCCAGTGGCCCAAGAACTGGTACA 492
Db |||||
Qy 4000 AGACTGTGAGACTGCGTGTGACCCCACTACGAGCCAGTGGCCCAAGAACTGGTACA 4059
Db |||||
Qy 493 TCAGCAGAACCCCAAGCAGAGGATGCTGTGTTGGGAGAGATGACCGATGGAT 552
Db |||||
Qy 4060 TCAGCAGAACCCCAAGCAGAGGATGCTGTGTTGGGAGAGATGACCGATGGAT 4119
Db |||||
Qy 553 TCAGTTCGAGTATGGCGGCCAGGGCTCCGACCTCTGCGGATGGCCATCCAGCTGACCT 612
Db |||||
Qy 4120 TCAGTTCGAGTATGGCGGCCAGGGCTCCGACCTCTGCGGATGGCCATCCAGCTGACCT 4179
Db |||||
Qy 613 TCCTGCGCCTGATGTCACCGAGGCTCCAGAACATCACTACCTGCAAGAACAGCG 672
Db |||||
Qy 4180 TCCTGCGCCTGATGTCACCGAGGCTCCAGAACATCACTACCTGCAAGAACAGCG 4239
Db |||||
Qy 673 TGGCTACATGGACAGAGACTGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 732
Db |||||
Qy 4240 TGGCTACATGGACAGAGACTGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 4299
Db |||||
Qy 733 ACAGAGTTCAGATCCGCGCGAGGGCAACAGCGCTTCACTACAGGCTCACTGTGATG 792
Db |||||
Qy 4300 ACAGAGTTCAGATCCGCGCGAGGGCAACAGCGCTTCACTACAGGCTCACTGTGATG 4359
Db |||||
Qy 793 GCTCAGAGTTCACACCGAGGCTGGGGCAAGACAGATGATGATACAAACACCAAGT 852
Db |||||
Qy 4360 GCTCAGAGTTCACACCGAGGCTGGGGCAAGACAGATGATGATACAAACACCAAGT 4419
Db |||||
Qy 853 CTTCCCGCTGCCCATCATGATGTGGCCCTTGGAGCTGTGGTCCCGACAGCAAGT 912
Db |||||
Qy 4420 CTTCCCGCTGCCCATCATGATGTGGCCCTTGGAGCTGTGGTCCCGACAGCAAGT 4479
Db |||||
Qy 913 TCGGCTTCGAGTGGCCCTGCTGCTTCTGTTAACTCCCTCCATC 959
Db |||||
Qy 4480 TCGGCTTCGAGTGGCCCTGCTGCTTCTGTTAACTCCCTCCATC 4526
Db |||||

RESULT 12
AEA04387

ID AEA04387 standard; cDNA; 5921 BP.

XX AEA04387;

XX 28-JUL-2005 (first entry)

DE Human cDNA from gene overexpressed in cancer, COL1A1.

XX Tumor marker; ss; gene; colon tumor; cancer; cytostatic; neoplasm;
diagnostic; microarray; drug screening.

XX Homo sapiens.

PN WO2005044990-A2.

PD 19-MAY-2005.

XX 01-NOV-2004; 2004WO-US036404.

XX 04-NOV-2003; 2003US-00700439.

XX (FARB) BAYER HEALTHCARE LLC.

PA (MAYO-) MAYO FOUND MEDICAL EDUCATION & RES.

XX Burgess C, Myerow S, Thiagalingam A, Maimonis P, Molino G;

PI BURGART L, Boardman LA, Thibodeau S, Lewis M;

XX WPI; 2005-372198/39.

DR P-PSDB; AEA04480.

DR REFSEQ; NM_000088.

XX Detecting expression of one or more nucleic acid sequences in biological
sample, useful for detecting cancer, comprises detecting a change in the
expression level of one or more nucleic acid sequences relative to a
control expression level.

PS Claim 1; SEQ ID NO 32; 256pp; English.

XX The invention relates to detecting differential expression of one or more
nucleic acid sequences (appearing as AEA04356-AEA04448 in a biological
sample comprising obtaining the sample from a subject, and detecting a
change in the expression level of one or more nucleic acid sequences
relative to a control expression level of the nucleic acid sequences, is
new. Also included are detecting cancer (or a pre-malignant condition,
thereof) in a subject (comprising comparing the expression level of one
or more nucleic acid sequences in a biological sample from the subject
with a control expression level of the nucleic acid sequences, where a
change of at least two-fold in the expression level of the nucleic acid
sequences is indicative of cancer or pre-malignant condition), monitoring
the onset (or progression, or regression) of cancer (or a pre-malignant
condition) in a subject (by detecting in a biological sample of the
subject at a first point in time the expression of one or more nucleic
acid sequences, repeating the first step at a subsequent point in time
and comparing the expression level detected, where a change in the
expression level is indicative of progression of cancer or its pre-
malignant condition in the subject), determining prognosis for cancer or
its pre-malignant condition in a subject (comprising detecting in a
biological sample of the subject, the expression level of one or more
nucleic acid sequences, comparing the expression level with a reference
expression level of the nucleic acid sequences and evaluating the
prognosis of the subject based on the comparison), determining the
efficacy of a test compound for inhibiting cancer in a subject,
determining the efficacy of a therapy for inhibiting cancer in a subject,
selecting a composition for inhibiting cancer in a subject, inhibiting
cancer in a subject, a polypeptide encoded by the nucleic acids above
(appearing as AEA04449-AEA04541), an antibody that specifically binds to
the polypeptide sequence, and detecting in a biological sample the
presence of a polypeptide. The method is useful for detecting in a
differential expression of one or more nucleic acid sequences in a

CC biological sample, which is useful for detecting cancer (especially colon
CC cancer), monitoring the onset, progression, or regression of cancer or a
CC pre-malignant condition, or determining prognosis for cancer or its pre-
CC malignant condition in a subject, or for determining the efficacy of a
CC test compound for inhibiting cancer in a subject. The compound is useful
CC for inhibiting cancer in a subject. The antibodies may also be used to
CC treat cancer. The present sequence is a cDNA from a human gene over-
CC expressed in cancer samples.
XX
SQ Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;

Query Match 97.8%; Score 942.2; DB 14; Length 5921;
Best Local Similarity 99.7%; Pred. No. 1.4e-191;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 13 GATCTAACGGTCTCCCTGGGCCCCATTGGGCCCCCTGCTCTCGGGTCGACACTGGTGATG 72
Db |||
Qy 3580 GACTCAACGGTCTCCCTGGGCCCCATTGGGCCCCCTGCTCTCGGGTCGACACTGGTGATG 3639
Db |||
Qy 73 CTGCTCTCTGTTGGTCCCGCGCCCTCTCGACCTCTGCTGCTCCCTGGTCCCTCCAGCG 132
Db |||
Qy 3640 CTGCTCTCTGTTGGTCCCGCGCCCTCTCGACCTCTGCTGCTCCCTGGTCCCTCCAGCG 3699
Db |||
Qy 133 CTGCTTTTCGACTTCAGCTTCTCTGCCCCAGCACCTCAAGAGAAGGCTCACGATGGTGCC 192
Db |||
Qy 3700 CTGCTTTTCGACTTCAGCTTCTCTGCCCCAGCACCTCAAGAGAAGGCTCACGATGGTGCC 3759
Db |||
Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGTTGCTGACCTGACCTGAGGTGCACCA 252
Db |||
Qy 3760 GCTACTACCGGGCTGATGATGCCAATGTGTTGCTGACCTGACCTGAGGTGCACCA 3819
Db |||
Qy 253 CCCTCAAGAGCCTGAGCCAGCAGATCAGAACATCCGGAGCCAGAGGGAAGCGCAAGA 312
Db |||
Qy 3820 CCCTCAAGAGCCTGAGCCAGCAGATCAGAACATCCGGAGCCAGAGGGAAGCGCAAGA 3879
Db |||
Qy 313 ACCCGCCCGCACCTGCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGGT 372
Db |||
Qy 3880 ACCCGCCCGCACCTGCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGGT 3939
Db |||
Qy 373 ACTGGATGTACCCCAACCAAGGCTGCAACCTGATGTCATCAAGTCTTCTGCAACATGG 432
Db |||
Qy 3940 ACTGGATGTACCCCAACCAAGGCTGCAACCTGATGTCATCAAGTCTTCTGCAACATGG 3999
Db |||
Qy 433 AGACTGTGTAGACTGCTGCTGTACCCCACTCAGCCAGTGGCCAGCAAGACTGGTACA 492
Db |||
Qy 4000 AGACTGTGTAGACTGCTGCTGTACCCCACTCAGCCAGTGGCCAGCAAGACTGGTACA 4059
Db |||
Qy 493 TCAGCAAGAACCCCAAGGCAAGAGGCACTGTCTGGTTGGCGAGAGCATCACCGATGGAT 552
Db |||
Qy 4060 TCAGCAAGAACCCCAAGGCAAGAGGCACTGTCTGGTTGGCGAGAGCATCACCGATGGAT 4119
Db |||
Qy 553 TCAGTTTCGATGTGGCGGCGAGGCTTCGACCTTCGGATGCGGATGCGCATCAGCTGACCT 612
Db |||
Qy 4120 TCAGTTTCGATGTGGCGGCGAGGCTTCGACCTTCGGATGCGGATGCGCATCAGCTGACCT 4179
Db |||
Qy 613 TCCTGCGCCTGATGCTCCACGAGGCTCCAGAACATCACTACCTGCAAGAACAGCG 672
Db |||
Qy 4180 TCCTGCGCCTGATGCTCCACGAGGCTCCAGAACATCACTACCTGCAAGAACAGCG 4239
Db |||
Qy 673 TGGCTTACATGGACAGCAGATGGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 732
Db |||
Qy 4240 TGGCTTACATGGACAGCAGATGGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 4299
Db |||
Qy 733 ACAGATTCGATGTCGGCGGAGGCAAGCGGCTTCACTACAGGCTCACTGTGATG 792
Db |||
Qy 4300 ACAGATTCGATGTCGGCGGAGGCAAGCGGCTTCACTACAGGCTCACTGTGATG 4359
Db |||
Qy 793 GCTGACGAGTCAACCGGAGCCTGGGGCAAGCAGTGATTTGAATACAAACACCAAGT 852
Db |||
Qy 4360 GCTGACGAGTCAACCGGAGCCTGGGGCAAGCAGTGATTTGAATACAAACACCAAGT 4419
Db |||
Qy 853 CTTCCCGCCTGCCCATCATCATGATGGCCCTTGGACGTTGGTGGCCCGACAGCAAT 912
Db |||

Db 4420 CTCCCGCTGCCCCATCATGATGTGGCCCCCTTGGACCTGTGGTGGCCAGACCAAT 4479
Qy 913 TGGGCTTCGACGTTGGCCCTGCTGCTCTCTCTGTAAACTCCCTCCATC 959
Db |||
Db 4480 TGGGCTTCGACGTTGGCCCTGCTGCTCTCTGTAAACTCCCTCCATC 4526
Db |||

RESULT 13
AAF90491
ID AAF90491 standard; DNA; 6728 BP.
XX
AC AAF90491;
XX
DT 22-AUG-2001 (first entry)
XX
DE Human pro-alpha-1 chain of type I procollagen DNA.
XX
KW COL1A1 gene; collagen; procollagen; human; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
CDS 120..4514
FT /*tag= a
FT sig_peptide 120..185
FT /*tag= b
FT mat_peptide 186..4511
FT /*tag= c
XX
PN WO200144455-A2.
XX
PD 21-JUN-2001.
XX
PF 12-DEC-2000; 2000WO-GB004741.
XX
PR 15-DEC-1999; 99GB-00029487.
XX
PA (ASTR) ASTRAZENECA AB.
PA (ASTR) ASTRAZENECA UK LTD.
XX
PI Beri R;
XX
DR WPI; 2001-398145/42.
DR P-PSDB; AAB82454.
XX
PT Novel antisense DNA oligonucleotide useful for inhibiting the expression
PT of wild type COL1A1 gene, for treating, reducing the risk of, and
PT preventing collagen disorders.
XX
PS Claim 1; Page 11-20; 30pp; English.
XX

CC The present sequence is that of the human gene for the pro-alpha-1 chain
CC of type I procollagen. The present invention relates to antisense
CC oligonucleotides (ASOs) and their use in inhibiting expression of type I
CC procollagen. The ASOs comprise 18-25 nucleotides and are complementary to
CC a nucleotide sequence from position 750 to position 3900 of the present
CC sequence. They are capable of inhibiting the expression of the
CC polypeptide encoded by the present sequence in a cell that expresses it.
CC Preferably, the ASO is complementary to a nucleotide sequence from
CC position 750-900, 1200-1300, 1400-1500, 1450-1550, 1850-2000, 2500-2600,
CC 2850-2950 and 3800-3900 of the present sequence, and is especially one of
CC those ASOs given in AAF90492-503. The ASOs are used in a claimed method
CC of treating, or reducing a risk of, a collagen disorder. Such disorders
CC may include those caused by overproduction of collagen fibres, such as
CC liver cirrhosis, kidney, liver and heart fibrosis, scleroderma,
CC hypertrophic scars and keloids
XX
SQ Sequence 6728 BP; 1213 A; 2144 C; 1971 G; 1400 T; 0 U; 0 Other;

Query Match 97.8%; Score 942.2; DB 4; Length 6728;
Best Local Similarity 99.7%; Pred. No. 1.4e-191;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX (AVAL-) AVALON PHARM.
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX Claim 1; SEQ ID NO 1066; 44pp; English.
XX The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
XX Sequence 6728 BP; 1213 A; 2144 C; 1971 G; 1400 T; 0 U; 0 Other;

Query Match 97.8%; Score 942.2; DB 6; Length 6728;
Best Local Similarity 99.7%; Pred. No. 1.4e-191;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 13 GATCTAACGGTCTCCCTGGGCCCAATGGGGCCCCCTGGTCTCGCGTTCGCACTGGTGATG 72
DB 3580 GACTCAACGGTCTCCCTGGGCCCAATGGGGCCCCCTGGTCTCGCGTTCGCACTGGTGATG 3639
QY 73 CTGGTCTCTGTGGTCCCGGGCCCTCTCGACCTCTGGTCTGGTCCCTGGTCCCGAGCG 132
DB 3640 CTGGTCTCTGTGGTCCCGGGCCCTCTCGACCTCTGGTCTGGTCCCGAGCG 3699
QY 133 CTGGTCTCTCGACTTCAGCTTCTGTCGCCCGCCAGCACTCAAGAGAGGCTCAGCATGGTGCC 192
DB 3700 CTGGTCTCTCGACTTCAGCTTCTGTCGCCCGCCAGCACTCAAGAGAGGCTCAGCATGGTGCC 3759
QY 193 GCTACTACCGGGCTGATGTCCTAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 252
DB 3760 GCTACTACCGGGCTGATGTCCTAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 3819
QY 253 CCCTCAAGAGCCTGAGCCAGCAGATCGAGACATCCGGAGCCCGAGAGGAGCGGCAAGA 312
DB 3820 CCCTCAAGAGCCTGAGCCAGCAGATCGAGACATCCGGAGCCCGAGAGGAGCGGCAAGA 3879
QY 313 ACCCGCCCGCACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGGTGGAGGT 372
DB 3880 ACCCGCCCGCACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGGTGGAGGT 3939
QY 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGG 432
DB 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGG 3999
QY 433 AGACTGTGTAGACCTGGCGTGTA CCCCACCTCAGCCCCAGTGGGCCCGCAGAGAACTGGGTACA 492

Search completed: April 27, 2006, 19:33:59
Job time : 604.014 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 17:29:04 ; Search time 4901.64 Seconds
(without alignments)
9192.024 Million cell updates/sec

Title: US-10-677-877a-1
Perfect score: 963
Sequence: 1 agcttcacgaagatctaac.....gtaaacctccctccatctaga 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: gb_est1.*
 - 2: gb_est2.*
 - 3: gb_est3.*
 - 4: gb_hic.*
 - 5: gb_est4.*
 - 6: gb_est5.*
 - 7: gb_est6.*
 - 8: gb_est7.*
 - 9: gb_gsa1.*
 - 10: gb_gsa2.*
 - 11: gb_gsa3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	925.4	96.1	3492	10	AY414729	Homo sapi
2	871.2	90.5	3378	10	AY414730	Pan trogl
3	803	83.4	1275	6	CF111062	Shultzomi
4	803	83.4	1275	6	CF111063	Shultzomi
5	779.8	81.0	3490	10	AY414731	Mus muscu
6	755.6	78.5	867	8	DN103076	DN103076
7	755.4	78.4	837	8	DN525976	DN525976
8	746.8	77.5	764	7	CN394423	CN394423
9	742.6	77.1	769	6	CD607436	CD607436
10	738.6	76.7	762	7	CN342438	CN342438
11	737.2	76.6	904	8	DR768506	DR768506
12	722.8	75.1	791	7	CJ023328	CJ023328
13	719	74.7	796	6	CD607437	CD607437
14	711	73.8	764	6	CD607434	CD607434
15	710.8	73.8	786	7	CJ026934	CJ026934
16	705.4	73.3	881	3	BI758827	BI758827
17	701.4	72.8	775	5	BW965531	BW965531
18	701	72.6	714	7	CN342233	CN342233
19	699.2	72.6	882	2	BI182334	BI182334
20	696.2	72.3	701	7	CN342240	CN342240
21	696.2	72.3	1035	1	AL568796	AL568796
22	694.2	72.1	700	7	CN342548	CN342548

23	691.4	71.8	896	6	CB194117	AGENCOURT
24	688	71.4	717	7	CN342260	CN342260
25	680.8	70.7	659	7	CN394414	CN394414
26	680.6	70.7	694	7	CN342173	CN342173
27	680	70.6	820	3	BP172484	BP172484
28	678.8	70.5	686	7	CN394454	CN394454
29	675.6	70.2	686	7	CN394556	CN394556
30	674.8	70.1	874	3	BP172845	BP172845
31	673.6	69.9	813	7	CV547993	CV547993
32	673.2	69.9	688	3	BM722597	BM722597
33	673	69.9	800	3	BP172993	BP172993
34	671.6	69.7	890	6	CA976611	CA976611
35	671	69.7	696	8	DR432524	DR432524
36	666.4	69.2	868	6	CF618321	CF618321
37	664.8	69.0	682	8	DR432984	DR432984
38	664.4	69.0	667	3	BM721080	BM721080
39	661	68.6	721	1	AW577435	AW577435
40	660.6	68.6	667	7	CN394438	CN394438
41	659.8	68.5	881	7	CN161921	CN161921
42	659.6	68.5	729	6	CD607386	CD607386
43	659.2	68.5	675	7	CN342526	CN342526
44	658	68.3	671	8	DR433321	DR433321
45	655.6	68.1	663	7	CN342398	CN342398

ALIGNMENTS

RESULT 1
LOCUS AY414729 3492 bp DNA linear GSS 17-DEC-2003
DEFINITION Homo sapiens COL1A1 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY414729
VERSION AY414729.1 GI:39770691
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 3492)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 3492)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment
FEATURES
source Location/Qualifiers
1..3492
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
<1..>3492
/genes="COL1A1"
/locus_tag="HQM5321"

ORIGIN
Query Match 96.1%; Score 925.4; DB 10; Length 3492;
Best Local Similarity 99.4%; Pred. No. 3.1e-200;
Matches 929; Conservative 0; Mismatches 6; Indels 0; Gaps 0;


```

RESULT 4
CF111063
LOCUS
DEFINITION
Shultzomic04314 Rat lung airway and parenchyma cDNA libraries
Rattus norvegicus cDNA clone Contig3827 5', mRNA sequence.
CF111063
ACCESSION
CF111063.1 GI:33167555
VERSION
EST.
KEYWORDS
Rattus norvegicus (Norway rat)
SOURCE
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 1275)
Shultz, W.A., Zhang, L., Gu, Y.-Z., Baker, G.L., Fannuchi, M.V.,
Padua, A.M., Gurske, W.A., Morin, D., Penn, S.G., Jovanovich, S.B.,
Plopper, C.G. and Buckpitt, A.R.
Gene expression analysis in response to lung toxicants: I.
Sequencing and microarray development
Am. J. Respir. Cell Mol. Biol. 30 (3), 296-310 (2004)
12947022
Contact: Shultz MA
Dept. of Molecular Biosciences, School of Veterinary Medicine
University of California, Davis
1311 Haring Hall, One Shields Avenue, Davis, CA 95616, USA
Tel: 530 752 0793
Fax: 530 752 4698
Email: mashultz@ucdavis.edu
Average Phred score is 20 or better. All poor quality data (Phred <
20) and vector/linker sequence has been removed.
High quality sequence stop: 1275.
FEATURES
source
location/Qualifiers
1..1275
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="Contig3827"
/sex="male"
/tissue_type="airway or parenchyma"
/dev_stage="adult"
/clone_lib="Rat lung airway and parenchyma cDNA libraries"
/notes="Organ: lung; Vector: pGEM-11zf(-); Site 1: Eco RI;
Site 2: Not I; mRNA was isolated from microdissected rat
lung airways and parenchyma tissues."
ORIGIN
Query Match 83.4%; Score 803; DB 6; Length 1275;
Best Local Similarity 90.5%; Pred. No. 2.3e-172;
Matches 857; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 13 GATCTAAGGTCTCCCTGGCCCATTTGGGGCCCTCGTGGTCTCGGTCGCGTGGCACTGGTGATG 72
DB 116 GACTCAACGGTCTCCCTGGCCCATTTGGTCCCTCGTCTCGAGGTGCGCACTGGCGATA 175
QY 73 CTGGTCTCTGTGGTCCCGCCCTCTGGAGCTCTGTGCTCCCTCGTCCCTCCAGCG 132
DB 176 GTGGTCTCTGTGGTCCCGCCCTCTGGAGCTCTGTGCTCCCTCGTCCCTCCAGCG 235
QY 133 CTGGTTTCGACTTCAGCTTCTCTGCCCGCAGCCACTCAAGAGAGGCTCAGCATGGTGGCC 192
DB 236 GTGGTTATGACTTCAGCTTCTCTGCCCTCAGCCACTCAAGAGAGTCTCAGATGGTGGCC 295
QY 193 GCTACTACCGGGTGTATGATGCAATGTGGTTCGTGACCGTGCACCTCGAGGTGGACCA 252
DB 296 GTTACTACCGGGCGGATGATGCAACGTCGTGCGTGCACCTGACCTTGGTGGACACTA 355
QY 253 CCTCAAGAGCTTGAGCCAGCATCGAGACATCCGGAGCCCGAGAGGAGCGGCAAGA 312
DB 356 CCTCAAGAGCTTGAGCCAGCATCGAGACATCCGGAGCCCGTGGAGGCGCGCAAGA 415
QY 313 ACCCGCGCCGCACTGCGCGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGT 372

```

```

DB 416 ACCCGCGCCGCACTGCGCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGCGGAGAT 475
QY 373 ACTTGAATTCACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGG 432
DB 476 ACTGGATGACCCCTAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGG 535
QY 433 AGACTGGTGCAGACCTGCGGTGTATCCCCACATCAGCCCATGTCGCCCAAGAACTGGTACA 492
DB 536 AGACAGGTTCAGACCTGTGTGTTCCCACTCAGCCCTCTGTGCTCAGAAGAACTGGTACA 595
QY 493 TCAGCAAGAAACCCCAAGGACATGTCGTGGTTCGGCGAGAGCATGACCGATGGAT 552
DB 596 TCAGCCCAAAACCCCAAGGAGAGCATGTCTGTGGTTTGGAGAGAGCATGACCGATGGAT 655
QY 553 TCCAGTTCCAGTATGCGCGCCGAGGCTCCGACCTTCGCCGATGTGGCCATCCAGCTGACCT 612
DB 656 TCAGTTCCAGTATGGAAGCAAGGTTCCGATCTTCGCCGATGTGCTATCCAGCTGACCT 715
QY 613 TCTCGGCTGTATGTCCACCGAGGCTCCCAAGAAATCACTACCACTGCAAGAACAGCG 672
DB 716 TCTCGGCTGTATGTCCACCGAGGCTCCCAAGAAATCACTATCACTGCAAGAACAGCG 775
QY 673 TGCCTTACATGACACAGAGATGCGCAACCTCAAGAGGCCCTGCTCTCAAGGGTCCA 732
DB 776 TAGCCTTACATGACCAACAGACTGCGCAACCTCAAGAAAGTCCCTGCTCTCCAGGGTCCA 835
QY 733 ACAGATTCAGATCCGCGCGGAGGCAACAGCGCTTCACTACACGCTCAGCTGTGATG 792
DB 836 ACAGATTCAGATTCAGGGGCAAGGCAACAGTGCATTCACCTACACACGCTTGTGATG 895
QY 793 GTTGACAGAGTCAACACCGAGGCTGGGGCAAGACAGTGTGAATCAAAACACCAAGT 852
DB 896 GCTGACAGAGTCAACACCGGAACTTGGGGCAAGACAGTGTGAATCAAAACACCAAGT 955
QY 853 CTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 912
DB 956 CTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1015
QY 913 TGGGCTTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 959
DB 1016 TGGGAATGACATTTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1062
RESULT 5
AY414731
LOCUS
DEFINITION
Mus musculus COL1A1 gene, VIRTUAL TRANSCRIPT1, partial sequence,
Genomic survey sequence.
AY414731
ACCESSION
AY414731.1 GI:39770693
VERSION
GSS.
KEYWORDS
Mus musculus (house mouse)
SOURCE
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridea; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 3490)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
Science 302 (5652), 1960-1963 (2003)
14671302
JOURNAL
PUBMED
2 (bases 1 to 3490)
Clark, A.G., Gnanowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
Todd, M.A., Tanenbaum, D.M., Civello, D.R., Lu, F., Murphy, B.,
Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Shinsky, J.J.,
Adams, M.D. and Cargill, M.
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,

```


COMMENT	Rockville, MD 20850, USA									
	This sequence was made by sequencing genomic exons and ordering them based on alignment.									
	Location/Qualifiers									
	source									
FEATURES	1..3490									
	/organism="Mus musculus"									
	/mol_type="genomic DNA"									
	/db_xref="taxon:10090"									
gene	<1..>3490									
	/gene="COL1A1"									
	/locus_tag="HCM5321"									
	ORIGIN									
Query Match 81.0%; Score 779.8; DB 10; Length 3490;										
Best Local Similarity 89.6%; Pred. No. 5.1e-167;										
Matches 838; Conservative 0; Mismatches 97; Indels 0; Gaps 0;										
Qy	13	GATCTAACGGTCTCCCTGGGCCCAATTGGGGCCCCCTGGTCTCTGCGGTGCGACTGGTGATG	72							
Db	2556	GACTCAACGGTCTCCCTGGGCCCAATTGGTCCCCCTGGTCTCTCGANGTTCGCACTGGTGACA	2615							
Qy	73	CTGTCCTCTGTTGGTCCCCCGGGCCCTCCTGGAACCTCTCTGGTTCCTCCCTGGTCTCCCAAGC	132							
Db	2616	GGGGCCCTGTGGTTCCTGCTCCCGGGCCCTCTCTGACCCCTCTCTGACCTCCCAAGT	2675							
Qy	133	CTGGTTTCGACTTCAGCTTCTCTGCCCCCAGCCACTCTCAAGAGAGGCTCAGATGGTGGCC	192							
Db	2676	GGGGTTATGACTTCAGCTTCTCTGCTCAGCCACCTCTCAAGAGAACTCTCAAGATGGTGGCC	2735							
Qy	193	GCTACTACCGGGCTGATGATGCCAAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA	252							
Db	2736	GCTACTACCGGGCCGATGATGCTTAACTGGTTCGTGACCGTGACCTTGAGTGGACACCA	2795							
Qy	253	CCCTCAAGAGCTGAGCAGCAGATCAGAAATCCGGAGCCCAAGAGGGAAGCGCAAGA	312							
Db	2796	CCCTCAAGAGCCTGAGTCAGCAGATTGAGAAATCCGCAGCCCGAAGGCAGCGCAAGA	2855							
Qy	313	ACCCGCCCGACCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT	372							
Db	2856	ACCTGCGCCGACATGCCGGACCTCAAGATGTGCCACTCTGACTGGAAGCGGAGAGT	2915							
Qy	373	ACTGGATTGACCCCAACAAAGGCTGCAACCTGGATGTCATCAAAGTCTTCTGTCAAAATGG	432							
Db	2916	ACTGGATGACCTTAACCAAGGCTGCAACCTGGAGCCCATCAAGTCTATCTGCAACATGG	2975							
Qy	433	AGACTGTGTGACCTGTGCTGTACCCCACTCAGCCCCAGTGTGGCCCAAGAACTGGTATCA	492							
Db	2976	AGA CAGGTCTCAGACCTGTGTGTTCCTCTACTCAGCCGCTCTGTGCTCTCAGAAGAACTGGTATCA	3035							
Qy	493	TCAGCAAGAACCCCAAGGACCAAGAGGACATGTCTGGTTTCGGCGAGAGCATCAACCGATGGAT	552							
Db	3036	TCAGCCCGAAACCCCAAGAAAGAACGACGTCTGGTTTGGAGAGAGCATCAACCGATGGAT	3095							
Qy	553	TCAGATTTCGAGTATGGCGGGCCAGGGCTCCGACCCCTGCGGATGTGGCCCATCCAGCTGACCT	612							
Db	3096	TCCCGTTTCGAGTACGGAAGCGAGGGCTCCGACCCCGCGGATGTGCTATTCAGCTGACCT	3155							
Qy	613	TCCTGCGCCTGATGTCTCACCGAGGCCCTCCCAGAAATCACTACCTACCTGCAAGAACAGCG	672							
Db	3156	TCCTGCGCCTAAATGTCTCACCGAGGCCCTCCCAAGAAATCACTATCACTGCAAGAACAGCG	3215							
Qy	673	TGGCCTACATGGACCAAGACACTGGCAACTCTCAGAGAGGCCCTGCTCTCTCAGGGGCTCCA	732							
Db	3216	TAGCCTACATGGACCAAGACACTGGCAACTCTCAGAGAGGCCCTGCTCTCTCAGGGGATCCA	3275							
Qy	733	ACGAGATTCGAGATTCGGCGCGCAGGGCAACAGCCGCTTCACTCAGCGGTCACTGTCTGATG	792							
Db	3276	ACGAGATTCGAGCTCAGAGGCGGAGGCAACAGTCCGCTTCACTCAGCACCCCTTGTGGACG	3335							
Qy	793	GCTGCAACGAGTTCACACCGGAGCCTGGGGCAAGACAGATGTATGTAATCAAAACCAACCAAGT	852							
Db	3336	GCTGCAACGAGTTCACACCGGAACTGGGGCAAGACAGTCACTCGAATCAAAACCAACCAAGA	3395							

```
Db 241 ATACTGGATTGACCCCAACCAAGGCTGCAACCTGCGAGCCCATCAAAAGTCTTCTGCAACAT 300
Qy 431 GGAGACTGGTGGAGACTGCTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTA 490
Db 301 GGAGACAGGCGAGACTGCTGTACCCCACTCAGCCAGGCTGCCCAAGAAAGTGGTA 360
Qy 491 CATCAGCAAGAACCCCAAGGACCAAGAGGCTGCTGTCTGGCGAGAGCATGACCGATGG 550
Db 361 CATCAGCAAGAACCCCAAGGACCAAGAGGCTGCTGTCTGGCGAGAGCATGACCGAGG 420
Qy 551 ATTCCAGTTTCAGATAGTGGCGGCGAGGCTCCGACCTCTGCGCGATGTGGCCATCCAGCTGAC 610
Db 421 ATTCCAGTTTCAGATAGTGGCGGCGAGGCTCCGACCTCTGCTGCTGAGCTGGCCATCCAGCTGAC 480
Qy 611 CTTTCCTGGCGCTGATGCTCAACCGAGGCTCCAGAGACATCACCTACCTGCAAGAACAG 670
Db 481 CTTTCCTGGCGCTGATGCTCAACCGAGGCTCCAGAGACATCACCTACCTGCAAGAACAG 540
Qy 671 CGTGGGCTTACATGACACAGCAGACTGGCAACCTCAAGAGGCGCTGTCTCTCAAGGGCTC 730
Db 541 CGTGGGCTTACATGACACAGCAGACTGGCAACCTCAAGAGGCGCTGTCTCTCCAGGGCTC 600
Qy 731 CAACGAGATCAGATCGCGCGCGAGGCAACAGCGCTTCACTACAGCGTCACTGTGGA 790
Db 601 CAACGAGATCAGATCGCGCGCGAGGCAACAGCGCTTCACTACAGCGTCACTGTGGA 660
Qy 791 TGCTGCACAGTCAACCGGAGCCTGGGCGAGAGAGTGAATGAATACAAACCAACCAA 850
Db 661 CGGCTGCACAGTCAACCGGAGCCTGGGCGAGAGAGTGAATGAATACAAACCAACCAA 720
Qy 851 GTCTCCCGCTCGCCATCATGATGTGGCGCCCTTGGAGCTTGGTGGCGCCCGAGCAGGA 910
Db 721 GACTTCCCGCTCGCCATCATGATGTGGCGCCCTTGGAGCTTGGTGGCGCCCGAGCAGGA 780
Qy 911 ATTGGCTTCGAGTTGGCGCTGTGCTGCTTCTGTAACTCCCTCC 956
Db 781 ATTGGCTTCGAGTTGGCGCTGTGCTGCTTCTGTAACTCCCTCC 826

RESULT 7
DN525976 837 bp mRNA linear EST 11-MAR-2005
LOCUS 1270578 MARC 7BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION DN525976
ACCESSION DN525976
VERSION DN525976.1 GI:60971208
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
1 (bases 1 to 837)
Smith,T.P.L., Roberts,A.J., Echtenkamp,S.E., Chitko-McKown,C.G.,
Wray,J.E. and Keefe,J.W.
A second set of bovine ESTs from pooled-tissue normalized libraries
Unpublished (2003)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called with phred v0.020425.c and
trimmed with the aid of the trim_alt option. Vector identified with
cross_match v0.990329.
Plate: RLK8043 row: J column: 3
Seq primer: GTAATGCACTCACTATAGG.
Location/Qualifiers
1..837
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
```

FEATURES

```
source
RESULT 8
CN394423 764 bp mRNA linear EST 16-MAY-2004
LOCUS 17000455488558 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
DEFINITION
```

```
/lab_host="DH10B"
/clone_lib="MARC 7BOV"
/notes="Vector: pcDNA3.1; Site 1: EcoRI; Site 2: NotI;
Library made with RNA pooled from multiple tissues
including ovary, hindbrain, uterus, and day-30 whole
embryos."

ORIGIN
Query Match 78.4%; Score 755.4; DB 8; Length 837;
Best Local Similarity 94.5%; Pred. No. 15e-161; Indels 0; Gaps 0;
Matches 783; Conservative 0; Mismatches 46;

Qy 13 GATCTAAACGCTCTCCCTGGCCCCCATTTGGGCCCCCTCTGCTCTCGCGTGTGCACTGGTGATG 72
Db 9 GACTCAATGCTCTCCAGAGCCCCCATCGGTCCCCCTGGGCTCGAGGTGGCACTGGTGATG 68
Qy 73 CTGGTCTCTGTTGGTCCCCCGGCCCCCTCTGAGACCTCTGCTGCTCCCTCTGCTCCCGAGG 132
Db 69 CTGGTCTCTGTTGGTCTCCCGGCCCCCTCTGAGACCTCTGCTGCTCCCGAGGCTCTCCAGG 128
Qy 133 CTGGTCTTTCAGCTTTCAGCTTCTGCCCCCAGCACCTCAAGAGAAAGGCTCAGCATGGTGGCC 192
Db 129 GCGGCTACGACTTGAAGCTTCTGCCCCAGCCACTCAGAGAAAGGCTCAGCATGGTGGCC 188
Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGTGTCGTCACCGTGAACCTCGAGGTGGACACCA 252
Db 189 GCTACTACCGGGCTGATGATGCCAATGTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 248
Qy 253 CCCTAAGAGCTTGAGCCAGCAGATCGAGAACATCCGAGGCCCGAGAGGAGCGGCAAGA 312
Db 249 CCCTAAGAGCTTGAGCCAGCAGATCGAGAACATCCGAGGCCCGTGAAGGAGCGGCAAGA 308
Qy 313 ACCCGCCCGCACCTCCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 372
Db 309 ACCCGCCCGCACCTCCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGCGGAGAT 368
Qy 373 ACTGGATTTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 432
Db 369 ACTGGATTTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 428
Qy 433 AGACTGGTGGAGCTCGGTGTACCCCACTCAGGCCAGTGTGGCCCGAGAGAACTGGTACA 492
Db 429 AAACCGGTGAGACTCTGTATATACCCCACTCAGGCCAGCGTGGCCCGAGAGAACTGGGTATA 488
Qy 493 TCAGCAAGAACCCCAAGGCAAGAGCATGTCTGTGTTGGCGCAGAGAGCATGCCGATGGAT 552
Db 489 TCAGCAAGAACCCCAAGGAAAGAGGCAAGTCTGTGTTACGGCGAGAGCATGCCGCGGAT 548
Qy 553 TCAGGTTTCGAGTATGGCGCCAGGCTCCGACCTCCGACCTGCGGATGTGGCCATCCAGTACCT 612
Db 549 TCAGGTTTCGAGTATGGCGCCAGGCTCCGATCCTGCGGATGTGGCCATCCAGTACCT 608
Qy 613 TCTGGCGCTGATGTCCAGAGGCTTCCAGAACATCACTTACCCTGCAAGAACAGCG 672
Db 609 TCTGGCGCTGATGTCCAGAGGCTTCCAGAACATCACTTACCCTGCAAGAACAGCG 668
Qy 673 TGGCTTACATGGACACGACACTGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 732
Db 669 TGGCTTACATGGACACGACACTGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 728
Qy 733 ACGAGATTCGAGATTCGCGCGCGAGGCAACAGCGCTTTCACCTACAGCGTCACTGTGATG 792
Db 729 ACGAGATTCGAGATTCGCGCGCGAGGCAACAGCGCTTTCACCTACAGCGTCACTGTGATG 788
Qy 793 GCTGCGAGCTCAACCGAGGCTGGGGGCAAGACAGTGAATGAATACAA 841
Db 789 GCTGCGAGCTCAACCGAGGCTGGGGGCAAGACAGTGAATGAATACAA 837

RESULT 8
CN394423
LOCUS
DEFINITION
```

```
ACCESSION      CN394423
VERSION        CN394423.1  GI:47382018
KEYWORDS       EST.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE      1 (bases 1 to 764)
AUTHORS        Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J.,
               Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R.,
               Lebkowski, J. and Stanton, L. W.
TITLE          Transcription characterization elucidates signaling networks that
               control human ES cell growth and differentiation
JOURNAL        Nat. Biotechnol. 22 (6), 707-716 (2004)
PUBMED         15146197
COMMENT        Contact: Brandenberger R
               Regenerative Medicine
               Genon Corporation
               230 Constitution Drive, Menlo Park, CA 94025, USA
               Tel: 650 473 8658
               Fax: 650 473 7760
               Email: rbrandenberger@genon.com
               Insert Length: 764 Std Error: 0.00.
               Location/Qualifiers
               1..764
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /tissue_type="embryonic stem cells, embryoid bodies
                  derived from H1, H7 and H9 cells"
                  /clone_lib="GN_EH"
                  /note="oligo dt primed, full-length enriched cDNA library
                  from embryoid body outgrowths derived from hES cell lines
                  H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
                  conditions."
FEATURES       source
               1..764
               77.5%; Score 746.8; DB 7; Length 764;
               Best Local Similarity 99.6%; Pred. No. 1.4e-159;
               Matches 759; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
               38 TGGGCCCCCTGGTCTCGCGTGGCACTGGTGATGCTGGTCTGTTGGTCCCCCGGCC 97
               Db
               3 TGGGCCCCCTGGTCTCGCGTGGCACTGGTGATGCTGGTCTGTTGGTCCCCCGGCC 62
               Qy
               98 TCCTGGACCTCCCTGGTCCCTGGTCTCCAGAGGCTGGTTTCGACTTCAGTTCCCTGCC 157
               Db
               63 TCCTGGACCTCCCTGGTCCCTGGTCTCCAGAGGCTGGTTTCGACTTCAGTTCCCTGCC 122
               Qy
               158 CCAGCCACCTCAAGAGAAGGCTCAGCATGGTGGCGCTACTACCGGCTGATGATGCCAA 217
               Db
               123 CCAGCCACCTCAAGAGAAGGCTCAGCATGGTGGCGCTACTACCGGCTGATGATGCCAA 182
               Qy
               218 TGTGGTTCGTGACCGTGACCTCGAGGTGGACACACACCTCAAGACCTGAGCCGAGCAT 277
               Db
               183 TGTGGTTCGTGACCGTGACCTCGAGGTGGACACACACCTCAAGACCTGAGCCGAGCAT 242
               Qy
               278 CGAGAACATCCGAGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTGCGGTGACCT 337
               Db
               243 CGAGAACATCCGAGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTGCGGTGACCT 302
               Qy
               338 CAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTGATGATGACCCCAACCAAGGCTG 397
               Db
               303 CAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTGATGATGACCCCAACCAAGGCTG 362
               Qy
               398 CAACCTGGATGCCATCAAAAGTCTTCTGGAACATGGAGACTGGTGAGACTGCGTGATACC 457
               Db
               363 CAACCTGGATGCCATCAAAAGTCTTCTGGAACATGGAGACTGGTGAGACTGCGTGATACC 422
               Qy
               458 CACTCAGCCAGTGTGGCCAGAGAACTGTTACATCAGCAAGAACCCCAAGGACAGAG 517
               Db
               423 CACTCAGCCAGTGTGGCCAGAGAACTGTTACATCAGCAAGAACCCCAAGGACAGAG 482
```

```
Qy 518 GCATGCTCTGTTGGCGGAGAGCATGACCGATGATTCAGTTCCAGTATGCGGCCGAGGG 577
Db 483 GCATGCTCTGTTGGCGGAGAGCATGACCGATGATTCAGTTCCAGTATGCGGCCGAGGG 542
Qy 578 CTCCGACCCCTGCCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTCCACCGAGGC 637
Db 543 CTCCGACCCCTGCCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTCCACCGAGGC 602
Qy 638 CTCCAGAAATCACCCTACCTACCTGCAAGAACAGCGTGGCTTACATGACACGACGACTGG 697
Db 603 CTCCAGAAATCACCCTACCTACCTGCAAGAACAGCGTGGCTTACATGACACGACGACTGG 662
Qy 698 CAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCAACGAG-ATCGAGATCGCGCCGAGG 756
Db 663 CAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCAACGAG-ATCGAGATCGCGCCGAGG 722
Qy 757 GCAACAGCGCTTTCACCTCAGCGTCACTGTCGATGGCTGCA 798
Db 723 GCAACAGCGCTTTCACCTCAGCGTCACTGTCGATGGCTGCA 764

RESULT 9
CD607436
LOCUS         CD607436             769 bp      mRNA      linear      EST 12-JAN-2004
DEFINITION   S6095953H1 FLP Homo sapiens cDNA, mRNA sequence.
ACCESSION    CD607436
VERSION      CD607436.1  GI:40255699
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
               Homnidae; Homo.
REFERENCE     1 (bases 1 to 769)
AUTHORS       Fu, G. K., Wang, J. T., Yang, J., Au-Young, J. and Stuve, L. L.
TITLE         Circular rapid amplification of cDNA ends for high-throughput
               extension cloning of partial genes
JOURNAL       Genomics 84 (1), 205-210 (2004)
PUBMED        15203218
COMMENT       Contact: Fu GK
               Incyte Genomics, Inc.
               3160 Porter Dr., Palo Alto, CA 94304, USA
               Tel: 6508454102
               Email: gfu@incyte.com.
               Location/Qualifiers
               1..769
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone_lib="FLP"
                  /note="Vector: pDrive Cloning Vector"
FEATURES      source
               1..769
               77.1%; Score 742.6; DB 6; Length 769;
               Best Local Similarity 99.3%; Pred. No. 1.3e-158;
               Matches 756; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
               89 CCGCGGCCCTCCCTGGACCTCTCGTCCCTCGTCTCCAGGCGCTGGTTTCGACTTCAG 148
               Db
               6 CCTGGCCCTTCCTGGACCTCTCGTCCCTCGTCTCCAGGCGCTGGTTTCGACTTCAG 65
               Qy
               149 CTTCCTGCCCCCAGCCACCTCAAGAGAAGGCTCAGATGGTGGCGCTACTACCGGGCTGA 208
               Db
               66 CTTCCTGCCCCCAGCCACCTCAAGAGAAGGCTCAGATGGTGGCGCTACTACCGGGCTGA 125
               Qy
               209 TGATGCCAATGTGGTTCGTGACCGTGCCTCGAGGTGGACACCACTCTCAAGACCTGAG 268
               Db
               126 TGATGCCAATGTGGTTCGTGACCGTGCCTCGAGGTGGACACCACTCTCAAGACCTGAG 185
               Qy
               269 CCAGCAGATCGAGAACATCCGGAGGCCAGAGGAGCCGCAAGAACCCCGCCGACCTG 328
               Db
               186 CCAGCAGATCGAGAACATCCGGAGGCCAGAGGAGCCGCAAGAACCCCGCCGACCTG 245
```


SOURCE Macaca mulatta (rhesus monkey)
 ORGANISM Macaca mulatta
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
 Cercopithecoidea; Cercopithecinae; Macaca.

REFERENCE 1 (bases 1 to 904)
 AUTHORS Magness, C.L., Fellin, P.C., Thomas, M.J., Korth, M.J., Agv, M.B.,
 Prohl, S.C., Fitzgibbon, M., Scherer, C.A., Miner, D.G., Katze, M.G. and
 Iadonato, S.P.

TITLE Analysis of the Macaca mulatta transcriptome and the sequence
 divergence between Macaca and human

JOURNAL Genome Biol. 6 (7), R60 (2005)
 PUBMED 15998449

COMMENT Contact: C. Magness
 Illumigen Biosciences Inc.
 2203 Airport Way S, Suite 450, Seattle, WA 98134, USA
 Tel: 2063780400
 Fax: 2063780408
 Email: cmagness@illumigen.com

Sequenced on 2005.04.01. 592 Q20 bases. Library Preparation: Prof.
 Michael Katze Lab at University of Washington DNA Sequencing:
 Illumigen Biosciences Inc. For further information, see
<http://www.macaque.org>

PCR Primers
 FORWARD: CCTCTACTAAGGGAACAAA
 BACKWARD: CACTATAGGCGGATTTGGGTA
 Insert Length: 904 Std Error: 0.00
 Plate: CL000631 row: D column: 06
 Seq primer: CCCTCACTAAAGGGAACAAA
 POLYA=Yes.

FEATURES
 source Location/Qualifiers
 1..904
 /organism="Macaca mulatta"
 /mol_type="mRNA"
 /strain="Indian"
 /db_xref="taxon:9544"
 /clone="IBIUN:31396"
 /sex="female"
 /dev_stage="adult"
 /lab_host="Electromax DH10B"
 /clone_lib="Katze MM0V"
 /notes="Organ: ovary; Vector: pDONR 222; Site 1: BarG 1;
 Site 2: BarG 1; Created from CloneMiner cdna_library
 Construction kit (catalog #18249-029)"

ORIGIN
 Query Match 76.6%; Score 737.2; DB 8; Length 904;
 Best Local Similarity 96.3%; Pred. No. 2.2e-157;
 Matches 765; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

170 AGAGAAGGCTCAGCATGGTGGCCGCTACTACCGGGCTGATGCCAATGTGGTTCGTGA 229
 |||||
 Db 5 AGAGAAGGCTCAGCATGGTGGCCGCTACTACCGGGCTGATGCCAATGTGGTTCGTGA 64

230 CCGTGACCTCAGGTGACACCCCTCAAGAGCTGAGCCAGCATGAGAACATCCG 289
 |||||
 Db 65 CCGTGACCTCAGGTGACACCCCTCAAGAGCTGAGCCAGCATGAGAACATCCG 124

290 GAGCCAGAGGAGCGCGAAGAACCCCGCCGACCTGCGCTGACCTCAAGATGTGCCA 349
 |||||
 Db 125 GAGCCCGAGGGCAGCCGCGAAGAACCCCGCCGACCTGCGCGACCTCAAGATGTGCCA 184

350 CTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGC 409
 |||||
 Db 185 CTCTGACTGGAAGAGCGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGC 244

410 CATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTGGTGATGCCCACTCAGGCCAG 469
 |||||
 Db 245 CATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTGGTGATGCCCACTCAGGCCAG 304

470 TGTGGCCCGAGAGAACCTGGTACATCAGCAAGAACCCCAAGGACCAAGAGGATCTCTGGTT 529
 |||||
 Db 305 TGTGGCCCGAGAGAACCTGGTACATCAGCAAGAACCCCAAGGACCAAGAGGACCTCTGGTT 364

QY 530 CGCGGAGAGCATGACCGATGCGATTCCAGTTTCGAGTATGGCGGCAGGGCTCCGACCTGC 589
 |||||
 Db 365 CGCGGAGAGCATGACCGACCGATTCCAGTTTCGAGTATGGCGGCAGGGCTCCGACCTGC 424
 |||||
 QY 590 CGATGTGGCCCATCCAGCTGACCTTCTTCGCGCTGATGTCACCGAGGCTCCCAAGAACAT 649
 |||||
 Db 425 CGAGCTGGCCCATCCAGCTGACCTTCTTCGCGCTGATGTCACCGAGGCTCCCAAGAACAT 484
 |||||
 QY 650 CACCTACCACTGCAAGAACAGCGTGCCTACATGGACCGACAGACTGGCAACCTCAAGAA 709
 |||||
 Db 485 CACCTACCACTGCAAGAACAGCGTGCCTACATGGACCGACAGACTGGCAACCTCAAGAA 544
 |||||
 QY 710 GSCCCTGCTCTCAAGGGCTCCAAACGAGATCGAGATCCGCGCGAGGGCAACAGCCGCTT 769
 |||||
 Db 545 GSCACTGCTCTCTCAAGGGCTCCAAACGAGATCGAGATCCGCGCGAGGGCAACAGTCGCTT 604
 |||||
 QY 770 CACTACAGGCTCACTGTGATGGCTGCAGAGTCAACCGAGCCCTGGGGCAAGACAGT 829
 |||||
 Db 605 CACTACAGGCTCACTGTGATGGCTGCAGAGTCAACCGAGCCCTGGGGCAAGACAGT 664
 |||||
 QY 830 GATTGAATACAAACCAACCAAGTCTCCCGCTGCCCATCATCGATGTGGCCCCCTTGA 889
 |||||
 Db 665 GATCGAATACAAACCAACCAAGTCTCCCGCTGCCCATCATCGATGTGGCCCCCTTGA 723
 |||||
 QY 890 CGTTGGTGGCCAGACAGGAATTGCGCTTCGAGCTTGGCCCTGTCTGCTTCTCTGTAAC 949
 |||||
 Db 724 CGTTGGTGGCCAGACAGGAATCCGCTTCCACGTTGGCCCGTCTGTTTCTTGGAAAC 783
 |||||
 QY 950 TCCCTCCACTCTAGA 963
 |||||
 Db 784 TCCCTCTCTCCACA 797

RESULT 12
 CJO23328
 LOCUS
 DEFINITION
 CJO23328 full-length enriched swine cdna library, adult trachea Sus
 scrofa cdna clone TCH01C120012 5', mRNA sequence.

ACCESSION CJO23328
 VERSION CJO23328.1 GI:54529668
 KEYWORDS EST.
 SOURCE Sus scrofa (pig)
 ORGANISM Sus scrofa

REFERENCE 1 (bases 1 to 791)
 AUTHORS Uenishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H.,
 Okumura, N., Hamasima, N. and Awata, T.

TITLE PEDE (Pig EST Data Explorer): construction of a database for ESTs
 derived from porcine full-length cdna libraries

JOURNAL Nucleic Acids Res. 32 (1), D484-D488 (2004)
 PUBMED 14681463

COMMENT Contact: Hirohide Uenishi
 Animal Genome Laboratory, Genome Research Department
 National Institute of Agrobiological Sciences
 2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
 Tel: +81-29-838-8627
 Fax: +81-29-838-8627
 Email: huenishi@affrc.go.jp

EST project with full-length enriched cdna libraries carried out in
 Animal Genome Research Program (Japan) by National Institute of
 Agrobiological Sciences and STAFF-Institute
 Single pass sequencing of clones derived from oligo-capped cdna
 library

Vector sequences were eliminated by RepeatMasker version 2002/07/13
 and crossmatch version 0.990319

Low quality bases were trimmed based on the quality values.
 FEATURES
 source Location/Qualifiers
 1..791
 /organism="Sus scrofa"
 /mol_type="mRNA"

```
/db_xref="taxon:9823"
/clone="TCH01C120012"
/tissue_type="trachea"
/dev_stage="adult"
/clone_lib="full-length enriched swine cDNA library, adult
trachea"

ORIGIN
Query Match 75.1%; Score 722.8; DB 7; Length 791;
Best Local Similarity 94.7%; Pred. No. 4.2e-154; Indels 0; Gaps 0;
Matches 748; Conservative 0; Mismatches 42;

Qy 154 TGCCCCAGCCACCTCAAGAGAGGCTCAGCATGCTGGCCGCTACTACCGGGCTGATGATG 213
Db 2 TGCCCCAGCCACCTCAAGAGAGGCTCAGCATGCTGGCCGCTACTACCGGGCCGATGATG 61

Qy 214 CCAATGTGGTCTGTGACCGTGACCTCGAGTGTGACACCAACCTCAAGAGGCTGAGCCAGC 273
Db 62 CCAATGTGGTCCGCGACCGTGACCTCGAGTGTGACACCAACCTCAAGAGGCTGAGCCAGC 121

Qy 274 AGATCGAGAACATCCGAGGCCAGAGGAGCCGCAAGAACCCGCCACCTGCCGTG 333
Db 122 AGATCGAGAACATCCGAGGCCCGAAGGCGAGCCGCAAGAACCCGCCACCTGCCGTG 181

Qy 334 ACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGTACTGTGATGACCCCAACCAAG 393
Db 182 ACCTCAAGATGTGCCACTCTGACTGGAAGAGCGGAGTACTGTGATGACCCCAACCAAG 241

Qy 394 GCTGCAACCTGAGATGCCATCAAGATCTTCTTGCAACATGGAGACTGTGAGACCTGGTGT 453
Db 242 GCTGCAACCTGAGACCGCGTCAAAAGTCTTCTTGCAACATGGAGAGCGGAGACCTGGTGT 301

Qy 454 ACCCACTCAGCCAGTGTGGCCAGAGAACTGTTACATCAGCAAGAACCCCAAGGACA 513
Db 302 ACCCACTCAGCCAGCGTCCCAAGAACTTGGTACATCAGCAAGAACCCCAAGGACA 361

Qy 514 AGAGGCATGTCTGTGGTGGCGAGAGCATGACCGATGGATTCAGTTCGAGTATGGCGCC 573
Db 362 AGAGGCATGTCTGTGGTGGCGAGAGCATGACCGAGGATTCAGTTCGAGTATGGCGCG 421

Qy 574 AGGGCTCGACCTTGGCGATGTGGCCATCCAGCTGACCTTCTTGGCGCTGATGTCAACCG 633
Db 422 AGGGCTCGGATCTGTGCTGAGTGGCCATCCAGCTGACCTTCTTGGCGCTGATGTCAACTG 481

Qy 634 AGGCCTCCAGAACATCACCTACCATGCAAGAACAGCGTGGCTACATGACGACAGCAGA 693
Db 482 AGGCCTCCAGAACATCACCTACCATGCAAGAACAGCGTGGCTACATGACGACAGCAGA 541

Qy 694 CTGGCAACCTCAAGAGGCGCTTCTCTCAAGGGCTCCAAACGAGATCGAGATCCGCGCCG 753
Db 542 CTGGCAACCTCAAGAGGCGCTTCTCTCAAGGGCTCCAAACGAGATCGAGATCCGCGCG 601

Qy 754 AGGGCAACAGCCGTTCACTAGAGTGTACTGTGATGCTGATGCTGACAGTTCACACCGAG 813
Db 602 AGGGCAACAGCCGTTCACTAGAGTGTACTGTGATGCTGATGCTGACAGTTCACACCGAG 661

Qy 814 CTTGGGCAAGACAGTGTGATTAACAACCAACCAAGTCTCCCGCTCCCATCATCG 873
Db 662 CTTGGGCAAGACAGTGTGATTAACAACCAACCAAGTCTCCCGCTCCCATCATCG 721

Qy 874 ATGTGGCCCCCTTGGAGTTGGTGGCCCGACAGCAAGAAATTCGGCTTTCGAGTTGGCCCTG 933
Db 722 ATGTGGCCCCCTTGGAGTTGGTGGCCCGCCGACCAAGAAATTCGGCAATTCGGCAATCG 781

Qy 934 TCTGCTTCCT 943
Db 782 TCTGCTTCCT 791

RESULT 13
CD607437/c 796 bp mRNA linear EST 12-JAN-2004
LOCUS CD607437
DEFINITION 56095953J1 FLP Homo sapiens cDNA, mRNA sequence.
```

```
ACCESSION CD607437 GI:40255700
VERSION CD607437.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Homnidae; Homo.
REFERENCE 1 (bases 1 to 796)
AUTHORS Fu, G.K., Wang, J.T., Yang, J., Au-Young, J. and Stuve, L.L.
TITLE Circular rapid amplification of cDNA ends for high-throughput
JOURNAL extension cloning of partial genes
PUBMED Genomics 84 (1), 205-210 (2004)
COMMENT 15203218
Contact: Fu GK
Incyte Genomics, Inc.
3160 Porter Dr., Palo Alto, CA 94304, USA
Tel: 6508454102
Email: gfu@incyte.com.
Location/Qualifiers
source
1..796
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="FLP"
/note="Vector: pDrive Cloning Vector"

ORIGIN
Query Match 74.7%; Score 719; DB 6; Length 796;
Best Local Similarity 99.0%; Pred. No. 3.1e-153;
Matches 755; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

Qy 150 TTCTCTGCCCCCAGCCACCTCAAGAGAGGCTCAGATGTTGGCGCTACTACCGGGCTGAT 209
Db 776 TTCTCTGCCCCCAGCCACCTCAAGAGAGGCTCAGATGTTGGCGCTACTACCGGGCTGAT 718

Qy 210 GATGCCAATGTGGTTCGTGACCCGTGA - CTTGAGGTGGAC - ACCACCTCAAGAGCCTGA 267
Db 717 GATGCCAATGTGGTTCGTGACCCGTGACCTCGAGGTGGACAAACACCTCAAGAGCCTGA 658

Qy 268 GCGAGCAGATCCAGACATCCGAGGCCAGAGGGAGCGCAAGACCCGCCCGCACCT 327
Db 657 GCGAGCAGATCCAGACATCCGAGGCCAGAGGGAGCGCGCAAGACCCGCCCGCACCT 598

Qy 328 GCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCA 387
Db 597 GCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCA 538

Qy 388 ACCAAGGCTGCAACCTGGATGCGCATCAAAAGTCTTCTTGCAACATGGAGACTGGTGAGACCT 447
Db 537 ACCAAGGCTGCAACCTGGATGCGCATCAAAAGTCTTCTTGCAACATGGAGACTGGTGAGACCT 478

Qy 448 GGGTGTACCCCTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCA 507
Db 477 GGGTGTACCCCTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCA 418

Qy 508 AGGACAAGAGGAGTGTCTGGTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATG 567
Db 417 AGGACAAGAGGAGTGTCTGGTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATG 358

Qy 568 GCGGCCAGGGCTCCGACCTGCGCGATGTGGCGCATTCAGCTGACCTTCTTCCGCGCTGATGT 627
Db 357 GCGGCCAGGGCTCCGACCTGCGCGATGTGGCGCATTCAGCTGACCTTCTTCCGCGCTGATGT 298

Qy 628 CCACCGAGGCTCCAGAACATCAGTACCTGCAAGAACAGCGTGGCTTACATGACC 687
Db 297 CCACCGAGGCTCCAGAACATCAGTACCTGCAAGAACAGCGTGGCTTACATGACC 238

Qy 688 AGCAGACTGGCAACCTCAAGAGAGGCGCTTCTCTTCAAGGGCTCCACAGAGATCGAGATCC 747
Db 237 AGCAGACTGGCAACCTCAAGAGAGGCGCTTCTCTTCAAGGGCTCCACAGAGATCGAGATCC 178

Qy 748 GCGCCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGCTGACAGAGTGCAC 807
```



```

Db      177  GCGCGAGGGCAACAGCGCTTACCTACAGGTCACTGTGATGGTGCACGAGTCACA 118
Qy      808  CCGAGAGCTGGGGCAAGACAGTGAATTAATAACAAACACCAAGTCTCTCCGCTGGCCCA 867
Db      117  CCGAGAGCTGGGGCAAGACAGTGAATTAATAACAAACACCAAGACCTCCCGCTTGCCCA 58
Qy      868  TCATCGATGTGGCCCGCTTGCACCTTGGTGGCCCGCCAGACCCAGGA 910
Db      57  TCATCGATGTGGCCCGCTTGGACGTTGGTGGCCCGCCAGACCCAGGA 15

RESULT 14
CD607434/C
LOCUS      CD607434      764 bp      mRNA      linear      EST 12-JAN-2004
DEFINITION 56087938H1 FLP Homo sapiens cDNA, mRNA sequence.
ACCESSION  CD607434
VERSION    CD607434.1  GI:40255697
KEYWORDS   EST...
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominoidea; Homo.
REFERENCE  1 (bases 1 to 764)
            Fu,G.K., Wang,J.T., Yang,J., Au-Young,J. and Stuve,L.L.
            Circular rapid amplification of cDNA ends for high-throughput
            extension cloning of partial genes
            Genomics 84 (1), 205-210 (2004)
JOURNAL   PUBMED 15203218
COMMENT   Contact: Fu GK
            Incyte Genomics, Inc.
            3160 Porter Dr., Palo Alto, CA 94304, USA
            Tel: 6508454102
            Email: gfu@incyte.com.
FEATURES   Location/Qualifiers
            source
            1..764
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone_lib="FLP"
            /note="Vector: pDrive Cloning Vector"

ORIGIN
Query Match 73.8%; Score 711; DB 6; Length 764;
Best Local Similarity 98.9%; Pred. No. 2e-151;
Matches 747; Conservative 0; Mismatches 5; Indels 3; Gaps 3;

Qy      125  TCCAGCGCTGGTTTCGACTTCA-GCTTCTGCCCCAGCCACCTCAAGAGAGGCTCAAG 183
Db      764  TCCAGCGCTGGTTTCGACTTCAAGGCTTCTGCTGCCCCAGCCACCTCCAGAGAGGCTCAAG 705
Qy      184  ATGTGTGCGGCTACTACCGGCTCATGATGCCAATGTGTTTCGACCGTGAACCTCGAGG 243
Db      704  ATGTGT-GCCGCTACTACCGGCTCATGATGCCAATGTGTTTCGACCGTGAACCTCGAGG 646
Qy      244  TGGACACCACTTCAAGAGCCTTGAGCCAGCAGATCGAGAAATCCGGAGCCCGAGAGGAA 303
Db      645  TGGACACCACTTCAAGAGCCTTGAGCCAGCAGATCGAGAAATCCGGAGCCCGAGA-GGCA 587
Qy      304  GCCGCAAGAACCCCGCGGCACTTGCCTGACCTCAAGATGTGCCATCTTGATCGGAAGA 363
Db      586  GCCGCAAGAACCCCGCGCGGCACTTGCCTGACCTCAAGATGTGCCATCTTGATCGGAAGA 527
Qy      364  GTGAGAGTACTGATTTGACCCCAACCAAGGCTCAACCTGGATGCCATCAAGTCTTCT 423
Db      526  GTGAGAGTACTGATTTGACCCCAACCAAGGCTCAACCTGGATGCCATCAAGTCTTCT 467
Qy      424  GCAACATGGAGACTGTTGAGACTTGCCTGTATACCCCACTCAGCCCAAGTGTGGCCCAAGA 483
Db      466  GCAACATGGAGACTGTTGAGACTTGCCTGTATACCCCACTCAGCCCAAGTGTGGCCCAAGA 407
Qy      484  ACTGGTACATCAGCAAGAACCCCAAGGCAAGAGGCGATGTCTGTGTTCCGGGAGAGCATGA 543

```

```

Db      406  ACTGTTACATCAGCAAGAACCCCAAGGCAAGAGGCGATGTTCTGGTTCGGCGAGAGCATGA 347
Qy      544  CCGATGGAATTCAGTTCGAGTATGGCGGCCAGAGGCTCCGACCCCTGCGATGTGGCCATCC 603
Db      346  CCGATGGAATTCAGTTCGAGTATGGCGGCCAGAGGCTCCGACCCCTGCGATGTGGCCATCC 287
Qy      604  AGCTGACCTTCTTCCGCTCATGTCCACCGAGGCTCCAGAGACATCACCCTACCCTGCA 663
Db      286  AGCTGACCTTCTTCCGCTCATGTTCACCGAGGCTCCAGAGACATCACCCTACCCTGCA 227
Qy      664  AGAACAGCGTGGCTCATATGACAGCAGCAGTGGCAACCTCAAGAGGCCCTGCTCTCTCA 723
Db      226  AGAACAGCGTGGCTCATATGACAGCAGCAGTGGCAACCTCAAGAGGCCCTGCTCTCTCC 167
Qy      724  AGGCTCTCAACGAGATCGAGATCCGCGCCGAGGCAACAGCCGCTTCACTTACAGCGTCA 783
Db      166  AGGCTCTCAACGAGATCGAGATCCGCGCCGAGGCAACAGCCGCTTCACTTACAGCGTCA 107
Qy      784  CTGTGATGCTGCGCAGTCAACCGAGCTCACCGAGCCTGGGCAAGACAGTGAATTAACAAA 843
Db      106  CTGTGATGCTGCGCAGTCAACCGAGCTCACCGAGCCTGGGCAAGACAGTGAATTAACAAA 47
Qy      844  CCACCAAGTCTTCCGCTCCCATCATCGATGTG 878
Db      46  CCACCAAGTCTTCCGCTCCCATCATCGAGTGTG 12

RESULT 15
CJ026934
LOCUS      CJ026934      786 bp      mRNA      linear      EST 22-OCT-2004
DEFINITION CJ026934 full-length enriched swine cDNA library, adult trachea Sus
            scrofa cDNA clone TCH01E040063 5', mRNA sequence.
ACCESSION  CJ026934
VERSION    CJ026934.1  GI:54537487
KEYWORDS   EST.
SOURCE     Sus scrofa (pig)
            Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
            Sus.
REFERENCE  1 (bases 1 to 786)
            Uenishi,H., Eguchi,T., Suzuki,K., Sawazaki,T., Toki,D., Shinkai,H.,
            Okumura,N., Hamashima,N. and Awata,T.
            PEDE (Pig EST Data Explorer): construction of a database for ESTs
            derived from porcine full-length cDNA libraries
            Nucleic Acids Res. 32 (1), D484-D488 (2004)
JOURNAL   PUBMED 14681463
COMMENT   Contact: Hirohide Uenishi
            Animal Genome Laboratory, Genome Research Department
            National Institute of Agrobiological Sciences
            2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
            Tel: +81-29-838-8627
            Fax: +81-29-838-8627
            Email: huenishi@affrc.go.jp
            EST project with full-length enriched cDNA libraries carried out in
            Animal Genome Research Program (Japan) by National Institute of
            Agrobiological Sciences and STAFF-Institute
            Single pass sequencing of clones derived from oligo-capped cDNA
            library
            Vector sequences were eliminated by RepeatMasker version 2002/07/13
            and crossmatch version 0.990319
            Low quality bases were trimmed based on the quality values.
FEATURES   Location/Qualifiers
            source
            1..786
            /organism="Sus scrofa"
            /mol_type="mRNA"
            /db_xref="taxon:9823"
            /clone="TCH01E040063"
            /tissue_type="trachea"
            /dev_stage="adult"
            /clone_lib="full-length enriched swine cDNA library, adult
            trachea"

```

ORIGIN

```
Query Match      73.8%; Score 710.8; DB 7; Length 786;
Best Local Similarity 94.0%; Pred. No. 2.3e-151;
Matches 739; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 23 TCTCCCTGGCCCAATTCGGGCCCTCGGTCTCGCGTTCGCACCTGGTGATGTCGTCCCTGT 82
Db 1 TCTCCCGGGCCCAATCGGTCCCTCGGTCTCGGTTCGCACCTGGTGATGTCGTCCCTGT 60

Qy 83 TGGTCCCGGGCCCTCTCGAATCTCGTTCCTCGTTCCTCGTTCCTCGTTCCTCGTTCCTCGA 142
Db 61 TGGTTCCTCCCGGGCCCTCTCGAATCTCGTTCCTCGTTCCTCGTTCCTCGTTCCTCGA 120

Qy 143 CTTTCAGCTTCCTGCGCCAGCCACTCAAGAGAGGGCTCAGATGGTGGCGGCTACTACCG 202
Db 121 CTTTCAGCTTCCTGCGCCAGCCACTCAAGAGAGGGCTCAGATGGTGGCGGCTACTACCG 180

Qy 203 GGGTGTGATGATGCCAATGTGGTTCGTGACCGTGCCTCGAGGTGGACACCACTCTCAAGAG 262
Db 181 GGGCGATGATGCCAATGTGGTTCGTGACCGTGCCTCGAGGTGGACACCACTCTCAAGAG 240

Qy 263 CTTGAGCCAGCAGATCCAGAAATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCCG 322
Db 241 CCTGAGCCAGCAGATCCAGAAATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCAG 300

Qy 323 CACTGCGGTGACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGACTCTGGATTGA 382
Db 301 CACTGCGGTGACTCAAGATGTGCCACTCTGACTGGAAGAGGGAAGTACTGGATTGA 360

Qy 383 CCCCACCAAGAGGTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGA 442
Db 361 CCCCACCAAGAGGTGCAACCTGGAGGCCATCAAGTCTTCTGCAACATGGAGACAGCGGA 420

Qy 443 GACTGCGGTGTA CCCCACCTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAA 502
Db 421 GACTGCGGTGTA CCCCACCTCAGCCAGCTGCCCCAGAGAACTGGTACATCAGCAAGAA 480

Qy 503 CCCCAGGACAGAGGGCTGTCTGGTTCGGGAGAGCATGACCGATGGATTCCAGTTTGA 562
Db 481 CCCCAGGACAGAGGGCACCTCTGGTACGGGAGAGCATGACCGAOGGATTCAGTTTGA 540

Qy 563 GTATGGCGGCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTTGCGCCT 622
Db 541 GTACGGCGGCGAGGGCTCCGATCTGTGACTGTGGCCATCCAGCTGACCTTCTTGCGCCT 600

Qy 623 GATGTCCACCGAGGCTCCAGAAATCACTACCTGCAAGAACTGGTACATCAGCGCTTACAT 682
Db 601 GATGTCCACTGAGGCTTCCCAAGAAATCACTACCTGCAAGAACTGGTACATCAGCGCTTACAT 660

Qy 683 GGACCCAGACTGGCAACCTCAGAGAGGGCCCTGCTCCTCAAGGGCTCCAAAGGATCGA 742
Db 661 GGACCCAGACTGGCAACCTCAGAGAGGGCCCTGCTCCTCAAGGGCTCCAAAGGATCGA 720

Qy 743 GATCCGGCGGAGGGCAACAGCCCTTCACTACAGGGTCACTGTTCGATGGCTGCACGAG 802
Db 721 GATCCGGCGGAGGGCAACAGCCCTTCACTACAGGGTCACTGTTCGATGGCTGCACGAG 780

Qy 803 TCACAC 808
Db 781 CCACAC 786
```

Search completed: April 27, 2006, 22:40:08
Job time : 4903.64 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 18:34:23 ; Search time 202.152 Seconds
(without alignments)
8467.835 Million cell updates/sec

Title: US-10-677-877A-1
Perfect score: 963
Sequence: 1 aagcttagtaagatctaac.....gtaaacctccctccatctaga 963

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/1 COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5 COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/1/ina/H COMB.seq.*
6: /cgn2_6/ptodata/1/ina/PCUTS COMB.seq.*
7: /cgn2_6/ptodata/1/ina/PP COMB.seq.*
8: /cgn2_6/ptodata/1/ina/RE COMB.seq.*
9: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	930	96.6	4409	3	US-09-331-347C-22
2	547.4	56.8	590	3	US-09-404-879A-215
3	547.4	56.8	590	3	US-09-338-933-215
4	547.4	56.8	590	3	US-09-215-681-215
5	547.4	56.8	590	3	US-09-216-003A-215
6	547.4	56.8	590	3	US-09-667-857-215
7	547.4	56.8	590	3	US-10-198-053-215
8	547.4	56.8	590	3	US-09-827-271-215
9	522.6	54.3	587	3	US-09-404-879A-192
10	522.6	54.3	587	3	US-09-338-933-192
11	522.6	54.3	587	3	US-09-215-681-192
12	522.6	54.3	587	3	US-09-216-003A-192
13	522.6	54.3	587	3	US-09-667-857-192
14	522.6	54.3	587	3	US-10-198-053-192
15	522.6	54.3	587	3	US-09-827-271-192
16	501.4	52.1	594	3	US-09-404-879A-214
17	501.4	52.1	594	3	US-09-338-933-214
18	501.4	52.1	594	3	US-09-215-681-214
19	501.4	52.1	594	3	US-09-216-003A-214
20	501.4	52.1	594	3	US-09-667-857-214
21	501.4	52.1	594	3	US-10-198-053-214
22	501.4	52.1	594	3	US-09-827-271-214
23	378.6	39.3	704	3	US-08-896-164-49
24	372.4	38.7	544	3	US-09-404-879A-257

25	372.4	38.7	544	3	US-09-338-933-257	Sequence 257, App
26	372.4	38.7	544	3	US-09-215-681-257	Sequence 257, App
27	372.4	38.7	544	3	US-09-216-003A-257	Sequence 257, App
28	372.4	38.7	544	3	US-09-667-857-257	Sequence 257, App
29	372.4	38.7	544	3	US-10-198-053-257	Sequence 257, App
30	372.4	38.7	544	3	US-09-827-271-257	Sequence 257, App
c 31	363	37.7	418	3	US-09-404-879A-258	Sequence 258, App
c 32	363	37.7	418	3	US-09-338-933-258	Sequence 258, App
c 33	363	37.7	418	3	US-09-215-681-258	Sequence 258, App
c 34	363	37.7	418	3	US-09-216-003A-258	Sequence 258, App
c 35	363	37.7	418	3	US-09-667-857-258	Sequence 258, App
c 36	363	37.7	418	3	US-10-198-053-258	Sequence 258, App
c 37	363	37.7	418	3	US-09-827-271-258	Sequence 258, App
c 38	348.6	36.2	5418	3	US-09-949-016-2477	Sequence 2477, App
39	347	36.0	5086	3	US-09-949-016-11	Sequence 11, Appl
40	339	35.2	1608	3	US-09-029-348-19	Sequence 19, Appl
41	332.4	34.5	18609	3	US-08-943-731-1	Sequence 1, Appl
c 42	318.2	33.0	343	3	US-09-702-705-418	Sequence 418, App
c 43	318.2	33.0	343	3	US-09-736-457-418	Sequence 418, App
c 44	318.2	33.0	343	3	US-09-614-124B-418	Sequence 418, App
c 45	318.2	33.0	343	3	US-09-671-325-418	Sequence 418, App

ALIGNMENTS

RESULT 1

US-09-331-347C-22
; Sequence 22, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
; TITLE OF INVENTION: obtaining Such and Their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331,347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 22
; LENGTH: 4409
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-331-347C-22

Query Match		96.6%	Score 930;	DB 3;	Length 4409;
Best Local Similarity		99.5%	Pred. No. 4.3e-208;		
Matches 933;		Conservative	0;	Mismatches	5;
				Indels	0;
				Gaps	0;
Qy	13	GATCTAACGGTCTCTCCCTGGGCCCCCATTTGGGCCCCCTGGTCTCTCGGGTCGCACCTGGTGTG	72		
Db	3471	GACTCAACGGTCTCTCCCTGGGCCCCCATTTGGGCCCCCTGGTCTCTCGGGTCGCACCTGGTGTG	3530		
Qy	73	CTGGTCTCTGTGGTCTCCCGGCCCCCTCTCGGACCTCTCTGTGTCCTCCCTGGTCTCTCCAGCG	132		
Db	3531	CTGGTCTCTGTGGTCTCCCGGCCCCCTCTCGGACCTCTCTGTGTCCTCCCTGGTCTCTCCAGCG	3590		
Qy	133	CTGGTCTCGACTTCAGCTTCTCCCGGCCCCCTCAAGAGAGGCTCAGCATGTGCGCC	192		
Db	3591	CTGGTCTCGACTTCAGCTTCTCCCGGCCCCCTCAAGAGAGGCTCAGCATGTGCGCC	3650		
Qy	193	GCTACTACCGGGCTGATGATGCGCAATGTGTTGCTGACCGCTGACCTCGAGGTGGACACCA	252		
Db	3651	GCTACTACCGGGCTGATGATGCGCAATGTGTTGCTGACCGCTGACCTCGAGGTGGACACCA	3710		
Qy	253	CCCTCAAGAGCTTGAGCCAGCAGATCGAGAACATTCGAGAGCCAGAGGAGCCGCAAGA	312		
Db	3711	CCCTCAAGAGCTTGAGCCAGCAGATCGAGAACATTCGAGAGCCAGAGGAGCCGCAAGA	3770		
Qy	313	ACCCGCGCCGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGGTGGAGGT	372		
Db	3771	ACCCGCGCCGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGGTGGAGGT	3830		

```
QY 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 432
|
|
|
Db 3831 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 3890
|
|
|
QY 433 AGACTGTGTAGACTGCGGTGTACCCCACTCAGCCCCAGTGTGGCCCAAGAACTGTGTACA 492
|
|
|
Db 3891 AGACTGTGTAGACTGCGGTGTACCCCACTCAGCCCCAGTGTGGCCCAAGAACTGTGTACA 3950
|
|
|
QY 493 TCAGCAGAACCCCAAGGACAAAGAGGCATCTCTGTTTCGGCGAGAGCATCAGCGATGGAT 552
|
|
|
Db 3951 TCAGCAGAACCCCAAGGACAAAGAGGCATCTCTGTTTCGGCGAGAGCATCAGCGATGGAT 4010
|
|
|
QY 553 TCAGATTTCGAGTATGGCGGCTCCAGGGCTCCGACCTCGCCGATGTGGCCATCCAGTGAACCT 612
|
|
|
Db 4011 TCAGATTTCGAGTATGGCGGCTCCAGGGCTCCGACCTCGCCGATGTGGCCATCCAGTGAACCT 4070
|
|
|
QY 613 TCCTGCGCCTGATGTCCACCGAGGCTCCCGAAGCATCACTACCACTGCAAGAACAGCG 672
|
|
|
Db 4071 TCCTGCGCCTGATGTCCACCGAGGCTCCCGAAGCATCACTACCACTGCAAGAACAGCG 4130
|
|
|
QY 673 TGGCCTACATGGAACACAGACTGGCAACCTCAAGAGGCGCTCTCTCAAGGGCTCCA 732
|
|
|
Db 4131 TGGCCTACATGGAACACAGACTGGCAACCTCAAGAGGCGCTCTCTCAAGGGCTCCA 4190
|
|
|
QY 733 ACAGATTCGAGATCCGCGCGAGGGCAACAGCGCTTCACTACAGGCTCACTGTGATG 792
|
|
|
Db 4191 ACAGATTCGAGATCCGCGCGAGGGCAACAGCGCTTCACTACAGGCTCACTGTGATG 4250
|
|
|
QY 793 GCTGCAGAGTCAACACCGAGGCTGGGGCAAGACAGTGAATACAAAACCAACCAAGT 852
|
|
|
Db 4251 GCTGCAGAGTCAACACCGAGGCTGGGGCAAGACAGTGAATACAAAACCAACCAAGT 4310
|
|
|
QY 853 CTTCCCGCCTGCCATCATCGATGTGGCCCCCTTGGACGTTGGTCCCGACGACGGAAT 912
|
|
|
Db 4311 CTTCCCGCCTGCCATCATCGATGTGGCCCCCTTGGACGTTGGTCCCGACGACGGAAT 4370
|
|
|
QY 913 TCGGCTTCGAGTGTGGCCCTGTCTGCTTCTCTGTAACCT 950
|
|
|
Db 4371 TCGGCTTCGAGTGTGGCCCTGTCTGCTTCTCTGTAAGCT 4408
|
|
|
```

RESULT 2

```
US-09-404-879A-215
; Sequence 215, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(590)
; OTHER INFORMATION: n = A,T,C or G
US-09-404-879A-215
```

```
Query Match 56.8%; Score 547.4; DB 3; Length 590;
Best Local Similarity 98.9%; Pred. No. 8.9e-119;
Matches 551; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 52 CTCGCGGTGCGACTGGTGATGCTGGTCTCTGTTGGTCCCCCGCCCTCTCTGGACCTCCTG 111
|
|
|
```

```
Db 23 CTCGCGGTGCGACTGGTGATGCTGGTCTCTGTTGGTCCCCCGCCCTCTCTGGACCTCCTG 82
|
|
|
QY 112 GTCCCTCTGTCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTCTGCCCCCAGCCACCTCAAG 171
|
|
|
Db 83 GTCCCTCTGTCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTCTGCCCCCAGCCACCTCAAG 142
|
|
|
QY 172 AGAAGGCTCACGATGTGTGGCCGCTACTACCGGGCTGATGATGCAATGTGGTTCTGTGACC 231
|
|
|
Db 143 AGAAGGCTCACGATGTGTGGCCGCTACTACCGGGCTGATGATGCAATGTGGTTCTGTGACC 202
|
|
|
QY 232 GTGACCTTCGAGGTGGGACACACCCCTCAAGAGCTTGAGCCAGCAGATCGGAACATCCGGA 291
|
|
|
Db 203 GTGACCTTCGAGGTGGGACACACCCCTCAAGAGCTTGAGCCAGCAGATCGGAACATCCGGA 262
|
|
|
QY 292 GCCCAGAGGGAAGCCCGCAAGAACCCCGCCGCGACCTTGCCTGTGACCTCAAGATGTGCCACT 351
|
|
|
Db 263 GCCCAGAGGGAAGCCCGCAAGAACCCCGCCGCGACCTTGCCTGTGACCTCAAGATGTGCCACT 322
|
|
|
QY 352 CTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACTGGATGCCA 411
|
|
|
Db 323 CTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACTGGATGCCA 382
|
|
|
QY 412 TCAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTGCGTGTACCCCACTCAGGCCAGTG 471
|
|
|
Db 383 TCAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTGCGTGTACCCCACTCAGGCCAGTG 442
|
|
|
QY 472 TGGCCCAGAGAACTCGTATCATCAGCAAGAACCCCAAGGACCAAGAGGCATGTCTGGTTG 531
|
|
|
Db 443 TGGCCCAGAGAACTCGTATCATCAGCAAGAACCCCAAGGACCAAGAGGCATGTCTGGTTG 502
|
|
|
QY 532 GCAGAGCATGACCCGATGGATTCCAGTTCGAGTATGGCGGCCAGGCTCGGACCCCTGGCG 591
|
|
|
Db 503 GCAGAGCATGACCCGATGGATTCCAGTTCGAGTATGGCGGCCAGGCTCGGACCCCTGGCG 562
|
|
|
QY 592 ATGTGCGCATCCAGCTG 608
|
|
|
Db 563 ATGTGCGCATCCGCGCG 579
|
|
|
```

RESULT 3

```
US-09-338-933-215
; Sequence 215, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; TITLE OF INVENTION: OVARIAN CANCER
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(590)
; OTHER INFORMATION: n = A,T,C or G
US-09-338-933-215
```

```
Query Match 56.8%; Score 547.4; DB 3; Length 590;
Best Local Similarity 98.9%; Pred. No. 8.9e-119;
Matches 551; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
```

```
QY 52 CTCGCGGTGCGACTGGTGATGCTGGTCTCTGTTGGTCCCCCGCCCTCTCTGGACCTCCTG 111
|
|
|
Db 23 CTCGCGGTGCGACTGGTGATGCTGGTCTCTGTTGGTCCCCCGCCCTCTCTGGACCTCCTG 82
|
|
|
QY 112 GTCCCTCTGTCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTCTGCCCCCAGCCACCTCAAG 171
|
|
|
```

Db 83 GTCCCCCTGGTCTCTCCAGCGCTGGTTTCGACTTCAGTTCCTGCTCCCGCCAGCCACTCAAG 142
Qy 172 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACC 231
Db 143 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACC 202
Qy 232 GTGACCTCGAGGTGGACACACCTCTCAAGAGCTTGAGCCCTGAGCAGATCGAGAACATCCGGA 291
Db 203 GTGACCTCGAGGTGGACACACCTCTCAAGAGCTTGAGCCCTGAGCAGATCGAGAACATCCGGA 262
Qy 292 GCCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTGCCCTGACCTCAAGATGTGCCACT 351
Db 263 GCCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTGCCCTGACCTCAAGATGTGCCACT 322
Qy 352 CTGACTCGAAGAGTGGAGAGTACTGGATTGACCTCAACCAAGAGCTGCAACTCGATGCCA 411
Db 323 CTGACTCGAAGAGTGGAGAGTACTGGATTGACCTCAACCAAGAGCTGCAACTCGATGCCA 382
Qy 412 TCAAGTCTTCTGCAACATGGAGACTGGTGAGACTTGGTGATCCGCTGACCCCACTCAGCCAGTG 471
Db 383 TCAAGTCTTCTGCAACATGGAGACTGGTGAGACTTGGTGATCCGCTGACCCCACTCAGCCAGTG 442
Qy 472 TGGCCCAAGAGAACTGGTATCATCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGGTTGCG 531
Db 443 TGGCCCAAGAGAACTGGTATCATCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGGTTGCG 502
Qy 532 GCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGCGGCGCAGGGCTCCGACCTGCCG 591
Db 503 GCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGCGGCGCAGGGCTCCGACCTGCCG 562
Qy 592 ATGTGGCCATCCAGCTG 608
Db 563 ATGTGGACCTCCGCGCG 579

RESULT 4

US-09-215-681-215
; Sequence 215, Application US/09215681A
; Patent No. 6528253
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; FILE REFERENCE: 210121.463
; CURRENT APPLICATION NUMBER: US/09/215,681A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(590)
; OTHER INFORMATION: n = A,T,C or G

US-09-215-681-215

Query Match 56.8%; Score 547.4; DB 3; Length 590;
Best Local Similarity 98.9%; Pred. No. 8.9e-119;
Matches 551; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 52 CTCGCGGTGCGACTGGTGATGCTGCTGTGGTCCCGCCGCTCTGGACCTCCTG 111
Db 23 CTCGCGGTGCGACTGGTGATGCTGCTGTGGTCCCGCCGCTCTGGACCTCCTG 82
Qy 112 GTCCCCCTGGTCTCCAGCGCTGGTTTCGACTTCAGTTCCTGCCCCAGCCACTCAAG 171
Db 83 GTCCCCCTGGTCTCCAGCGCTGGTTTCGACTTCAGTTCCTGCCCCAGCCACTCAAG 142
Qy 172 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACC 231

Db 143 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACC 202
Qy 232 GTGACCTCGAGGTGGACACACCTCTCAAGAGCTTGAGCCCTGAGCAGATCGAGAACATCCGGA 291
Db 203 GTGACCTCGAGGTGGACACACCTCTCAAGAGCTTGAGCCCTGAGCAGATCGAGAACATCCGGA 262
Qy 292 GCCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTGCCCTGACCTCAAGATGTGCCACT 351
Db 263 GCCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTGCCCTGACCTCAAGATGTGCCACT 322
Qy 352 CTGACTCGAAGAGTGGAGAGTACTGGATTGACCTCAACCAAGAGCTGCAACTCGATGCCA 411
Db 323 CTGACTCGAAGAGTGGAGAGTACTGGATTGACCTCAACCAAGAGCTGCAACTCGATGCCA 382
Qy 412 TCAAGTCTTCTGCAACATGGAGACTGGTGAGACTTGGTGATCCGCTGACCCCACTCAGCCAGTG 471
Db 383 TCAAGTCTTCTGCAACATGGAGACTGGTGAGACTTGGTGATCCGCTGACCCCACTCAGCCAGTG 442
Qy 472 TGGCCCAAGAGAACTGGTATCATCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGGTTGCG 531
Db 443 TGGCCCAAGAGAACTGGTATCATCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGGTTGCG 502
Qy 532 GCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGCGGCGCAGGGCTCCGACCTGCCG 591
Db 503 GCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGCGGCGCAGGGCTCCGACCTGCCG 562
Qy 592 ATGTGGCCATCCAGCTG 608
Db 563 ATGTGGACCTCCGCGCG 579

RESULT 5

US-09-216-003A-215
; Sequence 215, Application US/09216003A
; Patent No. 6670463
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF OVARIAN CANCER
; FILE REFERENCE: 210121.462
; CURRENT APPLICATION NUMBER: US/09/216,003A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 215
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (8)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (9)
; OTHER INFORMATION: Where n is a, c, g or t

US-09-216-003A-215

Query Match 56.8%; Score 547.4; DB 3; Length 590;
Best Local Similarity 98.9%; Pred. No. 8.9e-119;
Matches 551; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 52 CTCGCGGTGCGACTGGTGATGCTGCTGTGGTCCCGCCGCTCTGGACCTCCTG 111
Db 23 CTCGCGGTGCGACTGGTGATGCTGCTGTGGTCCCGCCGCTCTGGACCTCCTG 82
Qy 112 GTCCCCCTGGTCTCCAGCGCTGGTTTCGACTTCAGTTCCTGCCCCAGCCACTCAAG 171
Db 83 GTCCCCCTGGTCTCCAGCGCTGGTTTCGACTTCAGTTCCTGCCCCAGCCACTCAAG 142
Qy 172 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACC 231

Qy	292	GCCAGAGGAGACGCGAAGAACCCCGCCCGCAGCTGCGGTGACCTCAAGATGTGCAC	351
Db	263	GCCCAAGGGCAGNCGCAGAACCCCGCCCGCAGCTGCGGTGACCTCAAGATGTGCAC	322
Qy	352	CTGACTGGAAGAGTGGAGGAGTACTGGATTGACCCCAACCAAGGCTGCAACTCGATGCCA	411
Db	323	CTGACTGGAAGAGTGGAGGAGTACTGGATTGACCCCAACCAAGGCTGCAACTCGATGCCA	381
Qy	412	TCAAGCTCTTCTGCAACATGGAGAGCTGGTGAGAGCTGCGTGTATACCCCACTCAGCCCAAGTG	471
Db	382	TCAAGCTCTTCTGCAACATGGAGAGCTGGTGAGAGCTGCGTGTATACCCCACTCAGCCCAAGTG	441
Qy	472	TGCCCCAGAGAAGAACTGGTACATCAGCAAGAAACCCCAAGGCAAGAGGCATGTCTGGTTGG	531
Db	442	TGCCCCAAGAAGAACTGGTACATCAGCAAGAAACCCCAAGGCAAGAGGCATGTCTGGTTGG	501
Qy	532	GCAGAGACATGACCGATGGATTCCAGTTTCGAGTATGGCGGCCAAGGGCTCCGACCTTCGGCG	591
Db	502	GCAGAGAACATGACCGATGGATTCCAGTTTCGAGTATGGCGGCCAAGGGCTCCGACCTTCGGCG	561
Qy	592	ATGTGGCCAT	601
Db	562	ATGGGACCT	571

```

RESULT 10
US-09-338-933-192
  ; Sequence 192, Application US/093389933
  ; Patent No. 6488931
  ; GENERAL INFORMATION:
  ; APPLICANT: Mitcham, Jennifer Lynn
  ; APPLICANT: King, Gordon E.
  ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
  ; TITLE OF INVENTION: OVARIAN CANCER
  ; FILE REFERENCE: 210121.462C1
  ; CURRENT APPLICATION NUMBER: US/09/338,933
  ; CURRENT FILING DATE: 1999-06-23
  ; NUMBER OF SEQ ID NOS: 312
  ; SOFTWARE: FastSeq for Windows Version 3.0
  ; SEQ ID NO 192
  ; LENGTH: 587
  ; TYPE: DNA
  ; ORGANISM: Homo sapien
  ; FEATURE:
  ; NAME/KEY: misc_feature
  ; LOCATION: (1)...(587)
  ; OTHER INFORMATION: n = A,T,C or G
US-09-338-933-192

```

Qy	352	CTGACTCGAGAGTGGAGAGTACTTGGATTGACCCCAACCAAGGCTGCAACCTTGGATGCCA	411
Db	323	CTGACTTGGAAAGTGGAGAGTACTTGGATTGACCCCAACCAACCAAGGCTGCAACCTTGGATGCCA	381
Qy	412	TCAAAGTCTTCTGCAACAATCGAGAGACTTGGTGGAGACCTGCGTGTACCCCACTCAGCCCCAGTG	471
Db	382	TCAAAGTCTTCTGCAACAATCGAGAGACTTGGTGGAGACCTGCGTGTACCCCACTCAGCCCCAGTG	441
Qy	472	TGGCCCGAGAGAATCTGTGTACATCAGCAAGAACCCCAAGGACAAGGGCATGTCTGTGTTGG	531
Db	442	TGGCCCGCAAAAGAACTTGGTGTACATCAGCAGAAACCCCAAGGACAAGGGCATGTCTGTGTTGG	501
Qy	532	GCGAGAGCATCACCGATGGATTTCAGATTTCGAGTATGGCGGCACAGGGCTCCGACCCCTGCCG	591
Db	502	GCGAGAACATGACCGATGGATTTCAGATTTCGAGTATGGCGGCACAGGGCTCCGACCCCTGCCG	561
Qy	592	ATGTGGCCAT 601	
Db	562	ATGGGGACCT 571	

```

RESULT 11
US-09-215-681-192
; Sequence 192, Application US/09215681A
; Patent No. 6528253
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF OVARIAN CANCER
; FILE REFERENCE: 210121.463
; CURRENT APPLICATION NUMBER: US/09/215,681A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 192
; LENGTH: 587
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(587)
; OTHER INFORMATION: n = A,T,C or G
US-09-215-681-192

```


QY 412 TCAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTGCGTGTACCCCACTCAGCCCAAGT 471
DB 382 TCAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTGCGTGTACCCCACTCAGCCCAAGT 441
QY 472 TGSCCCAGAGAACTGCTATCAGCAAGAACCCCAAGGACAAGAGGCAATGCTGGTTGCG 531
DB 442 TGSCCCCAAGAACTGCTATCAGCAAGAACCCCAAGGACAAGAGGCAATGCTGGTTGCG 501
QY 532 GCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGGCGCCAGGCGCTCCGACCCCTGCCG 591
DB 502 GCGAGAACATGACCGATGGATTCCAGTTCGAGTATGGCGCGGCGCTCCGACCCCTGCCG 561
QY 592 ATGCGCCAT 601
DB 562 ATGGGACCT 571

RESULT 14
US-10-198-053-192
; Sequence 192, Application US/10198053
; Patent No. 6859710
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Hill, Paul
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.462C9
; CURRENT APPLICATION NUMBER: US/10/198,053
; CURRENT FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 624
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 192
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 276
; OTHER INFORMATION: n = A,T,C or G
US-10-198-053-192

Query Match 54.3%; Score 522.6; DB 3; Length 587;
Best Local Similarity 98.0%; Pred. No. 5.7e-113;
Matches 539; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 52 CTCGGGTGCACTGGTGTGCTGCTGTTGGTCCCGCCCTCTGGACCTCCTG 111
DB 23 CTCGGGTGCACTGGTGTGCTGCTGTTGGTCCCGCCCTCTGGACCTCCTG 82
QY 112 GTCCCTCTGCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTGCCCCAGCCACCTCAAG 171
DB 83 GCGCCCTCTGCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTGCCCCAGCCACCTCAAG 142
QY 172 GTCCCTCTGCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTGCCCCAGCCACCTCAAG 171
DB 83 GCGCCCTCTGCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTGCCCCAGCCACCTCAAG 142
QY 172 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGTTGCGTACC 231
DB 143 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGTTGCGTACC 202
QY 172 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGTTGCGTACC 231
DB 143 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGTTGCGTACC 202
QY 232 GTGACCTTCAGAGTGGACACCACTTCAGAGCCCTGAGCCAGAGATCGAGAAATCCGGA 291
DB 203 GTGACCTTCAGAGTGGACACCACTTCAGAGCCCTGAGCCAGAGATCGAGAAATCCGGA 262
QY 292 GCGCAGGGAAGCGCAAGAACCCCGCCGCACTTCCGCTGACCTCAAGATGCGCACT 351
DB 203 GTGACCTTCAGAGTGGACACCACTTCAGAGCCCTGAGCCAGAGATCGAGAAATCCGGA 262
QY 292 GCGCAGGGAAGCGCAAGAACCCCGCCGCACTTCCGCTGACCTCAAGATGCGCACT 351
DB 263 GCGCAGGGAAGCGCAAGAACCCCGCCGCACTTCCGCTGACCTCAAGATGCGCACT 322
QY 352 CTGACTGGAAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGATGCCA 411
DB 323 CTGACTGGAAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGATGCCA 381
QY 412 TCAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTGCGTGTACCCCACTCAGCCCAAGT 471

DB 382 TCAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTGCGTGTACCCCACTCAGCCCAAGT 441
QY 472 TGSCCCAGAGAACTGCTATCAGCAAGAACCCCAAGGACAAGAGGCAATGCTGGTTGCG 531
DB 442 TGSCCCCAAGAACTGCTATCAGCAAGAACCCCAAGGACAAGAGGCAATGCTGGTTGCG 501
QY 532 GCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGGCGCCAGGCGCTCCGACCCCTGCCG 591
DB 502 GCGAGAACATGACCGATGGATTCCAGTTCGAGTATGGCGCGGCGCTCCGACCCCTGCCG 561
QY 592 ATGCGCCAT 601
DB 562 ATGGGACCT 571

RESULT 15
US-09-827-271-192
; Sequence 192, Application US/09827271
; Patent No. 6962980
; GENERAL INFORMATION:
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.462C6
; CURRENT APPLICATION NUMBER: US/09/827,271
; CURRENT FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 192
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(587)
; OTHER INFORMATION: n = A,T,C or G
US-09-827-271-192

Query Match 54.3%; Score 522.6; DB 3; Length 587;
Best Local Similarity 98.0%; Pred. No. 5.7e-113;
Matches 539; Conservative 0; Mismatches 10; Indels 1; Gaps 1;
QY 52 CTCGGGTGCACTGGTGTGCTGCTGTTGGTCCCGCCCTCTGGACCTCCTG 111
DB 23 CTCGGGTGCACTGGTGTGCTGCTGTTGGTCCCGCCCTCTGGACCTCCTG 82
QY 112 GTCCCTCTGCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTGCCCCAGCCACCTCAAG 171
DB 83 GCGCCCTCTGCTCTCCAGCGCTGGTTTCGACTTCAGCTTCTGCCCCAGCCACCTCAAG 142
QY 172 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGTTGCGTACC 231
DB 143 AGAAGGCTCAGATGGTGGCGGCTACTACCGGGCTGATGATGCAATGTGTTGCGTACC 202
QY 232 GTGACCTTCAGAGTGGACACCACTTCAGAGCCCTGAGCCAGAGATCGAGAAATCCGGA 291
DB 203 GTGACCTTCAGAGTGGACACCACTTCAGAGCCCTGAGCCAGAGATCGAGAAATCCGGA 262
QY 292 GCGCAGGGAAGCGCAAGAACCCCGCCGCACTTCCGCTGACCTCAAGATGCGCACT 351
DB 263 GCGCAGGGAAGCGCAAGAACCCCGCCGCACTTCCGCTGACCTCAAGATGCGCACT 322
QY 352 CTGACTGGAAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGATGCCA 411
DB 323 CTGACTGGAAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCAACCTGATGCCA 381
QY 412 TCAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTGCGTGTACCCCACTCAGCCCAAGT 471
DB 382 TCAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTGCGTGTACCCCACTCAGCCCAAGT 441
QY 472 TGSCCCAGAGAACTGCTATCAGCAAGAACCCCAAGGACAAGAGGCAATGCTGGTTGCG 531

Db 442 TGGCCCAAAGAACTGGTACATCAGCAAGAACCCCAAGGACAAAGACATGCTGTGTTGG 501
Qy 532 GCGAGAGCATGACCCGATGGATTCCAGTTCGAGTATGGCGGCCAGGGCTCCGACCCCTGCCG 591
Db 502 GCGAGAACATGACCCGATGGATTCCAGTTCGAGTATGGCGGCCAGGGCTCCGACCCCTGCCG 561
Qy 592 ATGTGGCCAT 601
Db 562 ATGGGGACCT 571

Search completed: April 27, 2006, 18:58:40
Job time : 203.152 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 18:53:19 ; Search time 870.81 Seconds
(without alignments)
9144.835 Million cell updates/sec

Title: US-10-677-877A-1
Perfect score: 963
Sequence: 1 aagcttacgtaagatctaac.....gtaaacctccctccatctaga 963

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA Main:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	963	100.0	963	9 US-10-677-877A-1	Sequence 1, Appli
2	956.2	99.3	2487	9 US-10-677-877A-5	Sequence 5, Appli
3	952	98.9	2139	9 US-10-677-877A-13	Sequence 13, Appli
4	951	98.8	1734	9 US-10-677-877A-9	Sequence 9, Appli
5	942.2	97.8	5921	6 US-10-177-293-66	Sequence 66, Appli
6	942.2	97.8	5921	6 US-10-301-822-29	Sequence 29, Appli
7	942.2	97.8	5921	7 US-10-764-425-18	Sequence 18, Appli
8	942.2	97.8	5921	7 US-10-734-564-12	Sequence 12, Appli
9	942.2	97.8	5921	8 US-10-788-792-23	Sequence 23, Appli
10	942.2	97.8	5921	8 US-10-723-860-2288	Sequence 2288, Ap
11	942.2	97.8	5921	9 US-10-956-157-340	Sequence 340, App
12	942.2	97.8	5921	9 US-10-287-436A-139	Sequence 139, App
13	942.2	97.8	6728	3 US-09-954-456-782	Sequence 782, App
14	942.2	97.8	6728	3 US-09-880-107-3946	Sequence 3946, Ap
15	942.2	97.8	6728	3 US-09-918-715-260	Sequence 260, App
16	942.2	97.8	6728	3 US-09-873-367C-521	Sequence 521, App
17	942.2	97.8	6728	3 US-09-873-367C-521	Sequence 1066, Ap
18	942.2	97.8	6728	5 US-10-060-036-136	Sequence 136, App
19	942.2	97.8	6728	5 US-10-171-311-35	Sequence 35, Appli
20	942.2	97.8	6728	5 US-10-149-352-1	Sequence 1, Appli
21	942.2	97.8	6728	6 US-10-177-293-64	Sequence 64, Appli
22	942.2	97.8	6728	6 US-10-301-822-27	Sequence 27, Appli
23	942.2	97.8	6728	8 US-10-474-794-260	Sequence 260, App

24	942.2	97.8	6728	9 US-10-843-641A-521	Sequence 521, App
25	942.2	97.8	6728	9 US-10-843-641A-1066	Sequence 1066, Ap
26	942.2	97.8	6728	9 US-10-843-641A-3809	Sequence 3809, Ap
27	942.2	97.8	6728	9 US-10-852-335A-63	Sequence 63, Appli
28	942.2	97.8	6728	9 US-10-979-159-260	Sequence 260, App
29	942.2	97.8	6728	9 US-10-287-436A-755	Sequence 755, App
30	937.4	97.3	4744	5 US-10-084-817-49	Sequence 49, Appli
31	937.4	97.3	4770	6 US-10-291-265-7	Sequence 7, Appli
32	937.4	97.3	4791	6 US-10-291-265-479	Sequence 479, App
33	937.4	97.3	6739	5 US-10-116-802-237	Sequence 237, App
34	937.4	97.3	6753	5 US-10-198-846-13252	Sequence 13252, A
35	937	97.3	2192	3 US-09-925-301-42	Sequence 42, Appli
36	935.4	97.1	3347	5 US-10-000-773A-16	Sequence 16, Appli
37	930	96.6	4409	5 US-10-216-705-22	Sequence 22, Appli
38	926.4	96.2	6769	5 US-10-198-846-13897	Sequence 13897, A
39	914.2	94.9	1684	8 US-10-723-860-4837	Sequence 4837, Ap
40	912	94.7	1746	5 US-10-198-846-13901	Sequence 13901, A
41	908.6	94.4	6730	8 US-10-723-860-6498	Sequence 6498, Ap
42	899.4	93.4	2650	5 US-10-106-698-2090	Sequence 2090, Ap
43	888.6	90.2	1508	7 US-10-488-091-38	Sequence 38, Appli
44	862.2	89.5	4748	6 US-10-402-089-1	Sequence 1, Appli
45	862.2	89.5	4748	7 US-10-402-072A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-10-677-877A-1
; Sequence 1, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 1
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(947)
US-10-677-877A-1

Query Match		100.0%;	Score 963;	DB 9;	Length 963;
Best Local Similarity		100.0%;	Pred. No. 7.2e-261;		
Matches 963;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AAGCTTACGTAAGATCTAAGCGTCTCCCTGGCCCGCCATTTGGGCCCCCTGGTCTCTCGCGGTC	60		
Db	1	AAGCTTACGTAAGATCTAAGCGTCTCCCTGGCCCGCCATTTGGGCCCCCTGGTCTCTCGCGGTC	60		
Qy	61	GCATGTGTGATGCTGGTCTCTTGGTCCCGCCCGCCCTCTCGACCTCTTGGTCCCCCTG	120		
Db	61	GCATGTGTGATGCTGGTCTCTTGGTCCCGCCCGCCCTCTCGACCTCTTGGTCCCCCTG	120		
Qy	121	GTCTCCCGCGGCTGGTTTCGACTTCGCTTCTGCCCGCCAGCCACCTCAAGAGAGGCTC	180		
Db	121	GTCTCCCGCGGCTGGTTTCGACTTCGCTTCTGCCCGCCAGCCACCTCAAGAGAGGCTC	180		
Qy	181	ACGATGTGCGCGCTACTACCGGGGTGATGATGCAATGTGTTGGTGGACCTGACCTCG	240		
Db	181	ACGATGTGCGCGCTACTACCGGGGTGATGATGCAATGTGTTGGTGGACCTGACCTCG	240		
Qy	241	AGGTGGACACCACTTCAAGAGGCTTGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGG	300		
Db	241	AGGTGGACACCACTTCAAGAGGCTTGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGG	300		
Qy	301	GAAGCCCAAGAAACCCCGCCGACCTGCGCGCTCAAGATGTGCCACTTCTGACTGGA	360		

Db 301 GAAGCGCAAGAAACCCCGCCGACCTGCGGTGACCTCAAGATGTGCACTCTGACTGGA 360
Qy 361 AGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCACCTGGATGCCATCAAGTCT 420
Db 361 AGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCACCTGGATGCCATCAAGTCT 420
Qy 421 TCTGCAACATGGAGACTGGTGAGACCTGCTGTACCCCACTCAGCCAGTGGGCCAGA 480
Db 421 TCTGCAACATGGAGACTGGTGAGACCTGCTGTACCCCACTCAGCCAGTGGGCCAGA 480
Qy 481 AGAACTGGTATCATCAGCAAGAACCCCAAGGACCAAGAGGCAATGTCTGGTTCGGGAGAGCA 540
Db 481 AGAACTGGTATCATCAGCAAGAAACCCCAAGGACCAAGAGGCAATGTCTGGTTCGGGAGAGCA 540
Qy 541 TGAACCGATGGATTCAGTTCGAGTATGGCGGCCAGGGCTCGGACCTCGCGATGTGCCCA 600
Db 541 TGAACCGATGGATTCAGTTCGAGTATGGCGGCCAGGGCTCGGACCTCGCGATGTGCCCA 600
Qy 601 TCCAGCTGACCTTCTCGCGCTGATGTCACACGAGGCTCCAGAACATCACTACCACT 660
Db 601 TCCAGCTGACCTTCTCGCGCTGATGTCACACGAGGCTCCAGAACATCACTACCACT 660
Qy 661 GCAAGAACAGCGTGGCCTACATGCAACAGCAGACTGGCAACCTCAAGAGGCCCTGTCTCC 720
Db 661 GCAAGAACAGCGTGGCCTACATGCAACAGCAGACTGGCAACCTCAAGAGGCCCTGTCTCC 720
Qy 721 TCAAGGGCTCAACGAGATCGAGTCCGCGCCGAGGGCAACAGCCGCTTCACTACAGCG 780
Db 721 TCAAGGGCTCAACGAGATCGAGTCCGCGCCGAGGGCAACAGCCGCTTCACTACAGCG 780
Qy 781 TCACTGTGATGGCTGACAGTGCACACGAGCTCACACGAGGCTTGGGGCAAGACAGTATTGAATACA 840
Db 781 TCACTGTGATGGCTGACAGTGCACACGAGCTCACACGAGGCTTGGGGCAAGACAGTATTGAATACA 840
Qy 841 AAACCAACAAAGTCTCCCGCTGCCCCATCATGATGTGGCCCCCTTGGACGTTGGTGGCC 900
Db 841 AAACCAACAAAGTCTCCCGCTGCCCCATCATGATGTGGCCCCCTTGGACGTTGGTGGCC 900
Qy 901 CAGACAGGAATTCGGCTTCGAGTGGGCCCTGTCTGCTTCTCTGTAACCTCCCTCCATCT 960
Db 901 CAGACAGGAATTCGGCTTCGAGTGGGCCCTGTCTGCTTCTCTGTAACCTCCCTCCATCT 960
Qy 961 AGA 963
Db 961 AGA 963

RESULT 2
US-10-677-877A-5
; Sequence 5, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; TITLE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 5
; LENGTH: 2487
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2471)
US-10-677-877A-5

Query Match 99.3%; Score 956.2; DB 9; Length 2487;
Best Local Similarity 99.7%; Pred. No. 7.1e-259;
Matches 950; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 3 GCTTACGTAAAGATCTAAACGGTCTCTCGCGCCCAATTTGGGGCCCTCGGTCTCTCGCGGTGCG 62
Db 1527 GGTTCGGAAGATCTAAACGGTCTCTCGCGCCCAATTTGGGGCCCTCGGTCTCTCGCGGTGCG 1586
Qy 63 ACTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
Db 1587 ACTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1646
Qy 123 CTTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 182
Db 1647 CTTCCAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1706
Qy 183 GATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 242
Db 1707 GATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1766
Qy 243 GTGGACACCAACCTTCAAGAGCTTACCGGGCTGATGATGATGATGATGATGATGATGATGATGAT 302
Db 1767 GTGGACACCAACCTTCAAGAGCTTACCGGGCTGATGATGATGATGATGATGATGATGATGATGAT 1826
Qy 303 AGCGGCAAGAAACCCCGCCGCACTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362
Db 1827 AGCGGCAAGAAACCCCGCCGCACTTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1886
Qy 363 AGTGTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 422
Db 1887 AGTGTGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1946
Qy 423 TGCACATGTGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 482
Db 1947 TGCACATGTGAGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2006
Qy 483 AACTGTGTACATCAGCAAGAAACCCCAAGGACCAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 542
Db 2007 AACTGTGTACATCAGCAAGAAACCCCAAGGACCAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2066
Qy 543 ACCGATGGATTCAGTTCGAGTATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCT 602
Db 2067 ACCGATGGATTCAGTTCGAGTATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCT 2126
Qy 603 CAGCTGACCTTCTCGCGCTGATGCTCAACGAGGCTTCCAGAGCAATCACTACCTACCTACCTACCT 662
Db 2127 CAGCTGACCTTCTCGCGCTGATGCTCAACGAGGCTTCCAGAGCAATCACTACCTACCTACCTACCT 2186
Qy 663 AGAACAGAGGCT 722
Db 2187 AGAACAGAGGCT 2246
Qy 723 AAGGGCTCCAAACGAGATCGAGATCCGCGCGGCGGCAACAGCGCTTCACTTACAGCGCTC 782
Db 2247 AAGGGCTCCAAACGAGATCGAGATCCGCGCGGCGGCAACAGCGCTTCACTTACAGCGCTC 2306
Qy 783 ACTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 842
Db 2307 ACTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2366
Qy 843 ACCACCAAGTCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 902
Db 2367 ACCACCAAGTCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2426
Qy 903 GACACAGGAATTCGGCTTCCAGGCTTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 962
Db 2427 GACACAGGAATTCGGCTTCCAGGCTTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2486
Qy 963 A 963
Db 2487 A 2487

RESULT 3
US-10-677-877A-13
; Sequence 13, Application US/10677877A
; Publication No. US20050202537A1

GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 13
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (24)...(2123)
US-10-677-877A-13

Query Match 98.9%; Score 952; DB 9; Length 2139;
Best Local Similarity 100.0%; Pred. No. 1e-257;
Matches 952; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 AGATCTAACGGTCTCCCTGGCCCAATTGGGCCCTCTGGTCTCTCGGTGCGACTGGTGAT 71
Db AGATCTAACGGTCTCCCTGGCCCAATTGGGCCCTCTGGTCTCTCGGTGCGACTGGTGAT 1247

Qy 72 GCTGGTCTGTGGTCTCCCGGGCCCTCTGGACTCTCTGGTCTCTGGTCTCTCCAGC 131
Db GCTGGTCTGTGGTCTCCCGGGCCCTCTGGACTCTCTGGTCTCTGGTCTCTCCAGC 1307

Qy 132 GCTGGTTTTCAGTTTCACTTTCAGTCTCTGCCCCAGCCCTCAAGAGAGGCTCAAGTGGTGGC 191
Db GCTGGTTTTCAGTTTCACTTTCAGTCTCTGCCCCAGCCCTCAAGAGAGGCTCAAGTGGTGGC 1367

Qy 192 CGCTACTACCGGGCTGATGATGCCAATGTGTTCGTGACCGGTGACCTCGAGGTGGACAC 251
Db CGCTACTACCGGGCTGATGATGCCAATGTGTTCGTGACCGGTGACCTCGAGGTGGACAC 1427

Qy 252 ACCCTCAAGAGCTTGAGCCAGCAGATCGAGAACTCCGGAGCCCGAGAGGAGCGCGAAG 311
Db ACCCTCAAGAGCTTGAGCCAGCAGATCGAGAACTCCGGAGCCCGAGAGGAGCGCGAAG 1487

Qy 312 AACCCCGCCCGCACTGCGGTGACCTCAAGATGTGCCACTCTGACTCGAAGAGTGGAGAG 371
Db AACCCCGCCCGCACTGCGGTGACCTCAAGATGTGCCACTCTGACTCGAAGAGTGGAGAG 1547

Qy 372 TACTGGAATTGACCCCAACCAAGGCTGCAACTGTGATGCCATCAAAAGTCTTCTGCAACATG 431
Db TACTGGAATTGACCCCAACCAAGGCTGCAACTGTGATGCCATCAAAAGTCTTCTGCAACATG 1607

Qy 432 GAGACTGGTGAGACTTGGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTAC 491
Db GAGACTGGTGAGACTTGGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTAC 1667

Qy 492 ATCAGCAAGAACCCCAAGGACAGAGGATGTCTGTTTCGGCGAGAGCATGACCGATGGA 551
Db ATCAGCAAGAACCCCAAGGACAGAGGATGTCTGTTTCGGCGAGAGCATGACCGATGGA 1727

Qy 552 TTCCAGTTTCAGTATGCGGCGCAGGGCTCCGACCTGCCGATGTGGCCATCCAGCTGACC 611
Db TTCCAGTTTCAGTATGCGGCGCAGGGCTCCGACCTGCCGATGTGGCCATCCAGCTGACC 1787

Qy 612 TTCTGCGCTGATGTCCACCGAGGCTCCAGAACTCACTACCACTCAAGAGGCTCTCTCAAGAGGCTCC 671
Db TTCTGCGCTGATGTCCACCGAGGCTCCAGAACTCACTACCACTCAAGAGGCTCTCTCAAGAGGCTCC 1847

Qy 672 GTGGCTTACATGACCGAGCAGACTGGCAACTCAAGAGGCTCTCTCAAGAGGCTCTCTCAAGAGGCTCC 731
Db GTGGCTTACATGACCGAGCAGACTGGCAACTCAAGAGGCTCTCTCAAGAGGCTCTCTCAAGAGGCTCC 1907

Qy 732 AACGAGATCGAGATTCGCGCGGAGGGCAACAGCGCGTTTCACTTACAGCGTCACTGTTCGAT 791
Db AACGAGATCGAGATTCGCGCGGAGGGCAACAGCGCGTTTCACTTACAGCGTCACTGTTCGAT 1967

RESULT 4

US-10-677-877A-9
; Sequence 9, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 9
; LENGTH: 1734
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)...(1718)
US-10-677-877A-9

Query Match 98.8%; Score 951; DB 9; Length 1734;

Best Local Similarity 100.0%; Pred. No. 1.9e-257;
Matches 951; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 GATCTAACGGTCTCCCTGGCCCAATTGGGCCCTCTGGTCTCTCGGTGCGACTGGTGATG 72
Db GATCTAACGGTCTCCCTGGCCCAATTGGGCCCTCTGGTCTCTCGGTGCGACTGGTGATG 843

Qy 73 CTGGTCTGTGGTCTCCCGGGCCCTCTGGACTCTCTGGTCTCTGGTCTCTCCAGCG 132
Db CTGGTCTGTGGTCTCCCGGGCCCTCTGGACTCTCTGGTCTCTGGTCTCTCCAGCG 903

Qy 133 CTGGTTTTCAGTTTCACTTTCAGTCTCTGCCCCAGCCCTCAAGAGAGGCTCAAGTGGTGGCC 192
Db CTGGTTTTCAGTTTCACTTTCAGTCTCTGCCCCAGCCCTCAAGAGAGGCTCAAGTGGTGGCC 963

Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGTTCGTGACCGGTGACCTCGAGGTGGACACCA 252
Db GCTACTACCGGGCTGATGATGCCAATGTGTTCGTGACCGGTGACCTCGAGGTGGACACCA 1023

Qy 253 CCCTCAAGAGCTTGAGCCAGCAGATCGAGAACTCCGGAGCCCGAGAGGAGCGCGAAGA 312
Db CCCTCAAGAGCTTGAGCCAGCAGATCGAGAACTCCGGAGCCCGAGAGGAGCGCGAAGA 1083

Qy 313 ACCCCCGCCCGCACTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGGT 372
Db ACCCCCGCCCGCACTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGGT 1143

Qy 373 ACTGGATTGACCCCAACCAAGGCTGCCAACCCTGGATGCCATCAAAAGTCTTCTGCAACATGG 432
Db ACTGGATTGACCCCAACCAAGGCTGCCAACCCTGGATGCCATCAAAAGTCTTCTGCAACATGG 1203

Qy 433 AGACTGGTGAGACTTGGTGTACCCCACTCAGCCAGTGTGGCCCGAGAGAACTGGTACA 492
Db AGACTGGTGAGACTTGGTGTACCCCACTCAGCCAGTGTGGCCCGAGAGAACTGGTACA 1263

Qy 493 TCAGCAAGAACCCCAAGGACAGAGCATGTCTGTGGTTTCGGCGAGAGCATCAGCGATGGAT 552
Db TCAGCAAGAACCCCAAGGACAGAGCATGTCTGTGGTTTCGGCGAGAGCATCAGCGATGGAT 1323

```
QY 553 TCAGTTTCAGTATGGCGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCT 612
Db 1324 TCAGTTTCAGTATGGCGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCT 1383
QY 613 TCCTGGCCCTGATGTCCACGAGGCTCCAGAGGCTCCAGAACATCACTTACCACTGCAAGAACAGCG 672
Db 1384 TCCTGGCCCTGATGTCCACGAGGCTCCAGAGGCTCCAGAACATCACTTACCACTGCAAGAACAGCG 1443
QY 673 TGGCCCTACATGAGACAGACAGCTGGCAACCTCAAGAGGCCCCCTGCTCCCAAGGGCTCCA 732
Db 1444 TGGCCCTACATGAGACAGACAGCTGGCAACCTCAAGAGGCCCCCTGCTCCCAAGGGCTCCA 1503
QY 733 ACAGATTCAGAGTCCGCGCGAGGCAACAGACCGCTTCACTACAGCGTCACTGTGCGATG 792
Db 1504 ACAGATTCAGAGTCCGCGCGAGGCAACAGACCGCTTCACTACAGCGTCACTGTGCGATG 1563
QY 793 GCTGCAAGAGTCAACCGGAGCTGGGGCAAGACAGTGATTTGAATACAAACCAACAAGT 852
Db 1564 GCTGCAAGAGTCAACCGGAGCTGGGGCAAGACAGTGATTTGAATACAAACCAACAAGT 1623
QY 853 CTTCCCGCTTGGCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCAGACCAAGAT 912
Db 1624 CTTCCCGCTTGGCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCAGACCAAGAT 1683
QY 913 TCGGCTTCGACGTTGGCCCTGCTGCTTCTGTAAACTCCCTCCATCTAGA 963
Db 1684 TCGGCTTCGACGTTGGCCCTGCTGCTTCTGTAAACTCCCTCCATCTAGA 1734
```

RESULT 5

```
US-10-177-293-66
; Sequence 66, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 66
```

```
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-177-293-66

Query Match 97.8%; Score 942.2; DB 6; Length 5921;
Best Local Similarity 99.7%; Pred. No. 7.3e-255;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 13 GATCTAACGGTCTCCCTGGCCCCCATTTGGGCCCTTCTCGCGTTCGCACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGCCCCCATTTGGGCCCTTCTCGCGTTCGCACTGGTGATG 3639
QY 73 CTGGTCTCTTTGGTTCCTCCCGGCGCTCTCTGGACCTCTCTGGTTCCTTCCCGAGCG 132
Db 3640 CTGGTCTCTTTGGTTCCTCCCGGCGCTCTCTGGACCTCTCTGGTTCCTTCCCGAGCG 3699
QY 133 CTGGTTTCAGCTTCAGCTTCCTGCCCCAGCCACCTCAAGAGAGGCTCAGATGGTGGCC 192
Db 3700 CTGGTTTCAGCTTCAGCTTCCTGCCCCAGCCACCTCAAGAGAGGCTCAGATGGTGGCC 3759
QY 193 GCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACCGTGCACCTCGAGGTGGACACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACCGTGCACCTCGAGGTGGACACCA 3819
QY 253 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCGGAGCCCCAGAGGGAAGCCGAAGA 312
Db 3820 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCGGAGCCCCAGAGGGAAGCCGAAGA 3879
QY 313 ACCCGCCCGCACTGCGCTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAT 372
Db 3880 ACCCGCCCGCACTGCGCTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAT 3939
QY 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTCGATGCCATCAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTCGATGCCATCAAGTCTTCTGCAACATGG 3999
QY 433 AGACTGGTGCAGACCTGCGTGTATCCCCACTCAGCCCCAGTGTGGCCCCAGAGAACTGGTACA 492
Db 4000 AGACTGGTGCAGACCTGCGTGTATCCCCACTCAGCCCCAGTGTGGCCCCAGAGAACTGGTACA 4059
QY 493 TCAGCAAGAACCCCAAGGACAAGAGGATGTCTGGTTCGGCGAGAGCATGACCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGGACAAGAGGATGTCTGGTTCGGCGAGAGCATGACCGATGGAT 4119
QY 553 TCCAGTTTCAGTATGGCGGCCAGGGCTCCGACCCCTGCGCATGTGGCCATCCAGCTCACCT 612
Db 4120 TCCAGTTTCAGTATGGCGGCCAGGGCTCCGACCCCTGCGCATGTGGCCATCCAGCTCACCT 4179
QY 613 TCTGCGCTGATGTCCACCGAGGCTCCCGAGAACATCACTACCACTGCAAGAACAGCG 672
Db 4180 TCTGCGCTGATGTCCACCGAGGCTCCCGAGAACATCACTACCACTGCAAGAACAGCG 4239
QY 673 TGGCTTACATGAGCAGCAGATGCGCAACCTCAAGAGGCCCCCTGCTCCCAAGGGTCCA 732
Db 4240 TGGCTTACATGAGCAGCAGATGCGCAACCTCAAGAGGCCCCCTGCTCCCAAGGGTCCA 4299
QY 733 AGAGATTCAGATCCGCGCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATG 792
Db 4300 AGAGATTCAGATCCGCGCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATG 4359
QY 793 GCTGCAAGAGTCAACCGGAGCTTGGGGCAAGACAGTGAATGAATACAAACCAACAAGT 852
Db 4360 GCTGCAAGAGTCAACCGGAGCTTGGGGCAAGACAGTGAATGAATACAAACCAACAAGT 4419
QY 853 CTTCCCGCTTGGCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCAGACCAAGAT 912
Db 4420 CTTCCCGCTTGGCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCAGACCAAGAT 4479
QY 913 TCGGCTTCGACGTTGGCCCTGCTGCTTCTGTAAACTCCCTCCATC 959
Db 4480 TCGGCTTCGACGTTGGCCCTGCTGCTTCTGTAAACTCCCTCCATC 4526
```

[illegible]

```

RESULT 7
US-10-764-425-18
; Sequence 18, Application US/10764425
; Publication No. US20040146921A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; APPLICANT: Taylor, Ian
; TITLE OF INVENTION: EXPRESSION PROFILES FOR COLON CANCER AND METHODS OF USE
; FILE REFERENCE: 5151
; CURRENT APPLICATION NUMBER: US/10/764,425
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: 60/442,582
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-764-425-18

```

	Query Match	97.8%	Score 942.2	DB 7	Length 5921	
	Best Local Similarity	99.7%	Prod. No. 7.3e-255			
	Matches 944	Conservative 0	Mismatches 35	Indels 0	Gaps 0	
QY	13	GATCTAACGGCTCCCTCGGCCCAATTGGGCCCTCTGGTCTCGGGTTCGCACCTGGTGATG	72			
Db	3580	GACTCAACGGTCTCCCTGGGCCCAATTGGGCCCTCTGGTCTCGGGTTCGCACCTGGTGATG	363			
QY	73	CTGGTCTCTGTTGGTCCCCCGGGCCCTCTTGACCTCTCTGGTCCCCCTGGTCTCCCGAGG	132			
Db	3640	CTGGTCTGTTGGTCCCCCGGGCCCTCTTGACCTCTCTGGTCCCCCTGGTCTCCCGAGG	369			
QY	133	CTGGTTTCGACTTCAGCTTCTGCTCCCGACGACCTCAAGAAAGCTCACCAGTGGTGCC	192			
Db	3700	CTGGTTTCGACTTCAGCTTCTGCTCCCGACGACCTCAAGAAAGCTCACCAGTGGTGCC	375			
QY	193	GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA	252			

```

RESULT 6
US-10-301-822-29
; Sequence 29, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, Joan E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)...(4514)
US-10-301-822-29

```

Query Match	97.8%;	Score	942.2;	DB	6;	Length	5921;
Best Local Similarity	99.7%;	Pred. No.	7.3e-255;				
Matches	944;	Conservative	0;	Mismatches	3;	Indels	0;
QY	13	GATCTAACGGTCTCCCTGCGGCCCAATGCGGCCCTCTGCTGCTCGCGGTGCGACCTGGTGATG	72				
DB	3580	GACTCAACGGTCTCCCTGCGGCCCAATGCGGCCCTCTGCTGCTCGCGGTGCGACCTGGTGATG	3639				
QY	73	CTGGTCTCTGTTGTCCTCCCGGGCCCTCTCTGGACCTCTCTGGTCCCTCTGGTCTCTCCACAGC	132				
DB	3640	CTGGTCTCTGTTGTCCTCCCGGGCCCTCTCTGGACCTCTCTGGTCCCTCTGGTCTCTCCACAGC	3699				
QY	133	CTGCTTTTCGACTTCAGCTTCTCTGCCCCAGCCACTCTCAAGAGAAGGCTCAGCANGTGGCC	192				
DB	3700	CTGGTTTCGACTTCAGCTTCTCTGCCCCAGCCACTCTCAAGAGAAGGCTCAGCANGTGGCC	3759				
QY	193	GCTACTACCGGGCTGATGATGCCAATGTGTGTTCTGTGACCGTGAACCTCGAGGTGACACCA	252				
DB	3760	GCTACTACCGGGCTGATGATGCCAATGTGTGTTCTGTGACCGTGAACCTCGAGGTGACACCA	3819				
QY	253	CCCTCAGAGCCTGAGCCAGCAGATCGAGACATCCGGAGCCACAGAGGGAAGCCGCAAGA	312				
DB	3820	CCCTCAGAGCCTGAGCCAGCAGATCGAGAACATCCGGAGCCACAGAGGGAAGCCGCAAGA	3879				
QY	313	ACCCGGCCCGCACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT	372				
DB	3880	ACCCGGCCCGCACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT	3939				
QY	373	ACTGGATTGACCCCAACCAAGGCTGCACCTTGGATGCCATCAAAAGTCTTCTGCAACATGG	432				
DB	3940	ACTGGATTGACCCCAACCAAGGCTGCACCTTGGATGCCATCAAAAGTCTTCTGCAACATGG	3999				
QY	433	AGACTGGTGAGACCTGGGTGTACCCCACTCAGCCCACTGTGTGGCCCAAGACACTGGTACA	492				
DB	4000	AGACTGGTGAGACCTGGGTGTACCCCACTCAGCCCACTGTGTGGCCCAAGACACTGGTACA	4059				

```
Db 3760 GCTACTACCGGGCTGATGATGCCAATGTGGTTGGTGAACCTCGAGGTGACACCA 3819
Qy 253 CCCTCAAGAGCCTTGAGCCAGCAGATCGAGAACATCCGGAGCCACAGAGGGAAGCCGAAGA 312
Db 3820 CCCTCAAGAGCCTTGAGCCAGCAGATCGAGAACATCCGGAGCCACAGAGGGAAGCCGAAGA 3879
Qy 313 ACCCGCCGACACCTGCGCTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAT 372
Db 3880 ACCCGCCGACACCTGCGCTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGAGTGCATCAAACTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGAGTGCATCAAACTTCTGCAACATGG 3999
Qy 433 AGACTGTGAGACTCTGCGTGTACCCCACTCAGCCAGTGTGGCCGAGAACATGTGTACA 492
Db 4000 AGACTGTGAGACTCTGCGTGTACCCCACTCAGCCAGTGTGGCCGAGAACATGTGTACA 4059
Qy 493 TCAGCAAGAACCCCAAGGACCAAGAGGCAATGTCTGGTTTCGGCGAGAGCATCACCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGGACCAAGAGGCAATGTCTGGTTTCGGCGAGAGCATCACCGATGGAT 4119
Qy 553 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGAGTGTGGCCATCCAGCTGACCT 612
Db 4120 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGAGTGTGGCCATCCAGCTGACCT 4179
Qy 613 TCCTGGCCCTGATGTCCACCGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCG 672
Db 4180 TCCTGGCCCTGATGTCCACCGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCG 4239
Qy 673 TGGCTACATGAGACCAAGCAGATGGCAACCTCAAGAGGCGCTGCTCCTCAAGGGCTCCA 732
Db 4240 TGGCTACATGAGACCAAGCAGATGGCAACCTCAAGAGGCGCTGCTCCTCAAGGGCTCCA 4299
Qy 733 ACAGATTCGAGATCCGCGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATG 792
Db 4300 ACAGATTCGAGATCCGCGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATG 4359
Qy 793 GCTGCAGAGTCAACCGGAGCTTGGGGCAAGCAGTGTGAATACAAACCAACCAAGT 852
Db 4360 GCTGCAGAGTCAACCGGAGCTTGGGGCAAGCAGTGTGAATACAAACCAACCAAGT 4419
Qy 853 CCTCCCGCTGCCATCATCGATGTGGCCCTTGGACGTTGGTGGCCCAAGACAGCAAT 912
Db 4420 CCTCCCGCTGCCATCATCGATGTGGCCCTTGGACGTTGGTGGCCCAAGACAGCAAT 4479
Qy 913 TCGGCTTCGAGTGTGGCCCTGCTGCTTCTGTAAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTGTGGCCCTGCTGCTTCTGTAAACTCCCTCCATC 4526
```

RESULT 8

```
US-10-734-564-12
; Sequence 12, Application US/10734564
; Publication No. US20040157278A1
; GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using TIMP1
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10/734, 564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-734-564-12
```

Query Match 97.8%; Score 942.2; DB 7; Length 5921;
Best Local Similarity 99.7%; Pred. No. 7.3e-255;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy 13 GATCTAACGGTCTCCCTGGCCCCCATTGGGGCCCCCTGGTCTCTCGCGTTCGACTCGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGCCCCCATTGGGGCCCCCTGGTCTCTCGCGTTCGACTCGGTGATG 3639
Qy 73 CTGGTCTCTGTGGTCCCGCGCCCTCTCTGAGACTCTCTGGTCCCGCTCTCTCCAGCG 132
Db 3640 CTGGTCTCTGTGGTCCCGCGCCCTCTCTGAGACTCTCTGGTCCCGCTCTCTCCAGCG 3699
Qy 133 CTGGTTTCGACTTTCAGCTTCTCCGCCAGCACCTCAAGAGAGGCTCAGATGGTGGCC 192
Db 3700 CTGGTTTCGACTTTCAGCTTCTCCGCCAGCACCTCAAGAGAGGCTCAGATGGTGGCC 3759
Qy 193 GCTACTACCGGGCTGATGATGCAATGTGTGTCGTGACCGTGCCTCGAGGTGGACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCAATGTGTGTCGTGACCGTGCCTCGAGGTGGACCA 3819
Qy 253 CCCTCAAGAGCTTGAGCCAGCAGATCGAGAACATTCGCGAGCCACAGAGGGAAGCCGAGA 312
Db 3820 CCCTCAAGAGCTTGAGCCAGCAGATCGAGAACATTCGCGAGCCACAGAGGGAAGCCGAGA 3879
Qy 313 ACCCGCCGACCTCCCGTGCCTCAAGATGTGCCACTCTGACTGMAAGTGGAGAT 372
Db 3880 ACCCGCCGACCTCCCGTGCCTCAAGATGTGCCACTCTGACTGMAAGTGGAGAT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGAGTGCCTCAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGAGTGCCTCAAGTCTTCTGCAACATGG 3999
Qy 433 AGACTGTGAGACTCTGCGTGTACCCCACTCAGCCAGTGTGGCCCAAGAACATGTGTACA 492
Db 4000 AGACTGTGAGACTCTGCGTGTACCCCACTCAGCCAGTGTGGCCCAAGAACATGTGTACA 4059
Qy 493 TCAGCAAGAACCCCAAGGACCAAGAGCAGTGTCTGGTTTCGGCGAGAGCATCACCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGGACCAAGAGCAGTGTCTGGTTTCGGCGAGAGCATCACCGATGGAT 4119
Qy 553 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGAGTGTGGCCATCCAGCTGACCT 612
Db 4120 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGAGTGTGGCCATCCAGCTGACCT 4179
Qy 613 TCTCGCGCTGATGTCCACCGAGGCTTCCAGAACATCACTACCACTGCAAGAACAGCG 672
Db 4180 TCTCGCGCTGATGTCCACCGAGGCTTCCAGAACATCACTACCACTGCAAGAACAGCG 4239
Qy 673 TGGCTTACATGAGACCAAGCAGCAGATGGCAACCTCAAGAGGCGCTGCTCCTCAAGGGCTCCA 732
Db 4240 TGGCTTACATGAGACCAAGCAGCAGATGGCAACCTCAAGAGGCGCTGCTCCTCAAGGGCTCCA 4299
Qy 733 ACAGATTCGAGATCCGCGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATG 792
Db 4300 ACAGATTCGAGATCCGCGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATG 4359
Qy 793 GCTGCAGAGTCAACCGGAGCTTGGGGCAAGCAGTGTGAATACAAACCAACCAAGT 852
Db 4360 GCTGCAGAGTCAACCGGAGCTTGGGGCAAGCAGTGTGAATACAAACCAACCAAGT 4419
Qy 853 CCTCCCGCTGCCATCATCGATGTGGCCCTTGGACGTTGGTGGCCCAAGACAGCAAT 912
Db 4420 CCTCCCGCTGCCATCATCGATGTGGCCCTTGGACGTTGGTGGCCCAAGACAGCAAT 4479
Qy 913 TCGGCTTCGAGTGTGGCCCTGCTGCTTCTGTAAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTGTGGCCCTGCTGCTTCTGTAAACTCCCTCCATC 4526
```

RESULT 9

```
US-10-788-792-23
; Sequence 23, Application US/10788792
; Publication No. US20040191819A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
```


APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; CURRENT FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 23
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-788-792-23

Query Match 97.8%; Score 942.2; DB 8; Length 5921;
Best Local Similarity 99.7%; Pred. No. 7.3e-255;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

13 GATCTAAGGCTCTCCCTGGCCCAATGGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 72
Db |||||
3580 GACTCAACGGTCTCCCTGGCCCAATGGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 3639
Qy |||||
73 CTGGTCTGTGGTCCCGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 132
Db |||||
3640 CTGGTCTGTGGTCCCGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 3699
Qy |||||
133 CTGGTCTGTGGTCCCGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 192
Db |||||
3700 CTGGTCTGTGGTCCCGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 3759
Qy |||||
193 GCTACTACCGGGTGTGATGTCCTCAAGATGTGGTCTCGGGTGCACCTGGTATG 252
Db |||||
3760 GCTACTACCGGGTGTGATGTCCTCAAGATGTGGTCTCGGGTGCACCTGGTATG 3819
Qy |||||
253 CCCTCAAGAGCTGAGCCAGAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 312
Db |||||
3820 CCCTCAAGAGCTGAGCCAGAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 3879
Qy |||||
313 ACCCGCCCGCAGCTGCTGACCTCAAGATGTGGTCTCGGGTGCACCTGGTATG 372
Db |||||
3880 ACCCGCCCGCAGCTGCTGACCTCAAGATGTGGTCTCGGGTGCACCTGGTATG 3939
Qy |||||
373 ACTGGATTGACCTGAGCTGAGCCAGAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 432
Db |||||
3940 ACTGGATTGACCTGAGCTGAGCCAGAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 3999
Qy |||||
433 AGACTGGTGAAGCTCGGTGTACCCCACTCAGCCCGAGTGGCCCAAGAACTGGTATCA 492
Db |||||
4000 AGACTGGTGAAGCTCGGTGTACCCCACTCAGCCCGAGTGGCCCAAGAACTGGTATCA 4059
Qy |||||
493 TCAGCAAGAACCCCAAGAGCAAGAGGATGTCTGGTTCGGCGGAGAGCATACCGATGGAT 552
Db |||||
4060 TCAGCAAGAACCCCAAGAGCAAGAGGATGTCTGGTTCGGCGGAGAGCATACCGATGGAT 4119
Qy |||||

Db 4360 GCTGCAGAGTCAACCGGAGCTGGGCAAGACAGTATTGAATACAAAACCAACCAAGT 4419
Qy 853 CCTCCGCTGCGCCATCATCATGATGTCGCCCTTGGACGTTGGTGGCCCGACAGCAAGT 912
Db 4420 CCTCCGCTGCGCCATCATCATGATGTCGCCCTTGGACGTTGGTGGCCCGACAGCAAGT 4479
Qy 913 TCGGCTTCGAGCTTGGCCCTGCTCTGCTCTGTAAGTCTCCCTCCATC 959
Db 4480 TCGGCTTCGAGCTTGGCCCTGCTCTGCTCTGTAAGTCTCCCTCCATC 4526

RESULT 10
US-10-723-860-2288
; Sequence 2288, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Gineburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 2288
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-723-860-2288

Query Match 97.8%; Score 942.2; DB 8; Length 5921;
Best Local Similarity 99.7%; Pred. No. 7.3e-255;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

13 GATCTAAGGCTCTCCCTGGCCCAATGGGCCCTCTGGTCTCGGGTGCACCTGGTATG 72
Db 3580 GACTCAACGGTCTCCCTGGCCCAATGGGCCCTCTGGTCTCGGGTGCACCTGGTATG 3639
Qy 73 CTGGTCTGTGGTCCCGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 132
Db 3640 CTGGTCTGTGGTCCCGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 3699
Qy 133 CTGGTCTGTGGTCCCGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 192
Db 3700 CTGGTCTGTGGTCCCGGCCCTCTGGACCTCTGGTCTCGGGTGCACCTGGTATG 3759
Qy 193 GCTACTACCGGGTGTGATGTCCTCAAGATGTGGTCTCGGGTGCACCTGGTATG 252
Db 3760 GCTACTACCGGGTGTGATGTCCTCAAGATGTGGTCTCGGGTGCACCTGGTATG 3819
Qy 253 CCCTCAAGAGCTGAGCCAGAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 312
Db 3820 CCCTCAAGAGCTGAGCCAGAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 3879
Qy 313 ACCCGCCCGCAGCTGCTGACCTCAAGATGTGGTCTCGGGTGCACCTGGTATG 372
Db 3880 ACCCGCCCGCAGCTGCTGACCTCAAGATGTGGTCTCGGGTGCACCTGGTATG 3939
Qy 373 ACTGGATTGACCTGAGCTGAGCCAGAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 432
Db 3940 ACTGGATTGACCTGAGCTGAGCCAGAGATCGAGAACATCCCGAGCCCGAGAGGAGCGCAAGA 3999
Qy 433 AGACTGGTGAAGCTCGGTGTACCCCACTCAGCCCGAGTGGCCCAAGAACTGGTATCA 492
Db 4000 AGACTGGTGAAGCTCGGTGTACCCCACTCAGCCCGAGTGGCCCAAGAACTGGTATCA 4059
Qy 493 TCAGCAAGAACCCCAAGAGCAAGAGGATGTCTGGTTCGGCGGAGAGCATACCGATGGAT 552
Db 4060 TCAGCAAGAACCCCAAGAGCAAGAGGATGTCTGGTTCGGCGGAGAGCATACCGATGGAT 4119
Qy 4119


```
Db 3580 GACTCAACGGTCTCCCTGGCCCCATTGGGCCCCCTGGTCTCTCGCGTCCGACTGGTGATG 3639
Qy 73 CTGCTCTCTGTTGGTCCCCCGGCGCTCTGAGACTCTCTGGTCCCGCTGGTCTCTCCAGCG 132
Db 3640 CTGGTCTCTGTTGGTCCCCCGGCGCTCTCTGAGACTCTCTGGTCCCGCTGGTCTCTCCAGCG 3699
Qy 133 CTGGTTTCGACTTCAGCTTCTCTGCCCCAGCCACTCAAGAGAGGCTCACGATGGTGGCC 192
Db 3700 CTGGTTTCGACTTCAGCTTCTCTGCCCCAGCCACTCAAGAGAGGCTCACGATGGTGGCC 3759
Qy 193 GCTACTACCGGGCTGATGATCCAAATGTGGTTCGTGACCTCGAGTCCGAGGTGACACCA 252
Db 3760 GCTACTACCGGGCTGATGATCCAAATGTGGTTCGTGACCTCGAGTCCGAGGTGACACCA 3819
Qy 253 CCCTCAAGAGCTGAGCCAGCAGATCGAGAACTCCGGAGGCCAGAGAGGAGCCGCAAGA 312
Db 3820 CCCTCAAGAGCTGAGCCAGCAGATCGAGAACTCCGGAGGCCAGAGAGGAGCCGCAAGA 3879
Qy 313 ACCCGCCCGCACCTCGCGTCACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAT 372
Db 3880 ACCCGCCCGCACCTCGCGTCACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCTCAAAAGTCTTCTGCAACATGG 432
Db 4000 AGACTGTGAGACTCTGCGTGTACCCCTCAGCCCTCAGCCAGTGTGGCCAGAGAACTGGTACA 4059
Qy 493 TCAGCAAGAAACCCCAAGGACAGAGGCAATGTCTGGTTCGGCGAGAGCATGCCATGGAT 552
Db 4060 TCAGCAAGAAACCCCAAGGACAGAGGCAATGTCTGGTTCGGCGAGAGCATGCCATGGAT 4119
Qy 553 TCCAGTTTCGAGTATGGCGGCCAGAGGCTCCGACCTCGAGTGGCCATCAAAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCTCAAAAGTCTTCTGCAACATGG 3999
Qy 433 AGACTGTGAGACTCTGCGTGTACCCCTCAGCCCTCAGCCAGTGTGGCCAGAGAACTGGTACA 492
Db 4000 AGACTGTGAGACTCTGCGTGTACCCCTCAGCCCTCAGCCAGTGTGGCCAGAGAACTGGTACA 4059
Qy 493 TCAGCAAGAAACCCCAAGGACAGAGGCAATGTCTGGTTCGGCGAGAGCATGCCATGGAT 552
Db 4060 TCAGCAAGAAACCCCAAGGACAGAGGCAATGTCTGGTTCGGCGAGAGCATGCCATGGAT 4119
Qy 553 TCCAGTTTCGAGTATGGCGGCCAGAGGCTCCGACCTCGAGTGGCCATCAAAAGTCTTCTGCAACATGG 432
Db 4120 TCCAGTTTCGAGTATGGCGGCCAGAGGCTCCGACCTCGAGTGGCCATCAAAAGTCTTCTGCAACATGG 4179
Qy 613 TCCGCGCTGATGTCCACCGAGGCTCCCAAGAACTACCTACCTACCTGCAAGAAACAGCG 672
Db 4180 TCCGCGCTGATGTCCACCGAGGCTCCCAAGAACTACCTACCTACCTGCAAGAAACAGCG 4239
Qy 673 TGGCTTACATGGACAGCAGATCGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 732
Db 4240 TGGCTTACATGGACAGCAGATCGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 4299
Qy 733 ACAGATCGAGATCCGCGCCGAGGCAACAGCCGCTTCACTACAGGCTCACTGTGATG 792
Db 4300 ACAGATCGAGATCCGCGCCGAGGCAACAGCCGCTTCACTACAGGCTCACTGTGATG 4359
Qy 793 GCTGACGAGTTCACACCGGAGCTGGGGCAAGACAGATGATGATGATGATGATGATGATG 852
Db 4360 GCTGACGAGTTCACACCGGAGCTGGGGCAAGACAGATGATGATGATGATGATGATG 4419
Qy 853 CTTCCCGCTGCCATCATCATGATGTGGCCCTTTGGAGCTGTGGTGGCCCGACAGCAAGAT 912
Db 4420 CTTCCCGCTGCCATCATCATGATGTGGCCCTTTGGAGCTGTGGTGGCCCGACAGCAAGAT 4479
Qy 913 TCGGCTTCGAGTGTGGCCCTGTCTGCTTCCGTAACTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTGTGGCCCTGTCTGCTTCCGTAACTCCCTCCATC 4526
```

RESULT 13

```
US-09-954-456-782
; Sequence 782, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
; TITLE OF INVENTION: Setb
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
```

```
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233, 617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234, 052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234, 923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235, 134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235, 637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235, 638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235, 711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235, 863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 782
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-954-456-782
```

```
Query Match 97.8%; Score 942.2; DB 3; Length 6728;
Best Local Similarity 99.7%; Pred. No. 7.5e-255;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy 13 GATCTAACGGTCTCCCTGGCCCCCATTTGGGCCCCCTGGTCTCTCGCGTCCGACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGCCCCCATTTGGGCCCCCTGGTCTCTCGCGTCCGACTGGTGATG 3639
Qy 73 CTGGTCTCTGTTGGTCCCCCGGCGCTCTCTGGAACCTCTCTGTCCTCTGTCCTCTGTCCTCCAGCG 132
Db 3640 CTGGTCTCTGTTGGTCCCCCGGCGCTCTCTGGAACCTCTCTGTCCTCTGTCCTCTCCAGCG 3699
Qy 133 CTGGTTTCGACTTCAGCTTCTCTGCCCCAGCCACTCAAGAGAGGCTCACGATGGTGGCC 192
Db 3700 CTGGTTTCGACTTCAGCTTCTCTGCCCCAGCCACTCAAGAGAGGCTCACGATGGTGGCC 3759
Qy 193 GCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACCTCGAGTCCGAGGTGACACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACCTCGAGTCCGAGGTGACACCA 3819
Qy 253 CCCTCAAGAGCTGAGCCAGCAGATCGAGAACTCCGGAGGCCAGAGGAGGAGCCGCAAGA 312
Db 3820 CCCTCAAGAGCTGAGCCAGCAGATCGAGAACTCCGGAGGCCAGAGGAGGAGCCGCAAGA 3879
Qy 313 ACCCGCCCGCACCTCGCGTCACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAT 372
Db 3880 ACCCGCCCGCACCTCGCGTCACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCTCAAAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCTCAAAAGTCTTCTGCAACATGG 3999
Qy 433 AGACTGTGAGACTCTGCGTGTACCCCTCAGCCCTCAGCCAGTGTGGCCAGAGAACTGGTACA 492
Db 4000 AGACTGTGAGACTCTGCGTGTACCCCTCAGCCCTCAGCCAGTGTGGCCAGAGAACTGGTACA 4059
Qy 493 TCAGCAAGAAACCCCAAGGACAGAGGCAATGTCTGGTTCGGCGAGAGCATGCCATGGAT 552
Db 4060 TCAGCAAGAAACCCCAAGGACAGAGGCAATGTCTGGTTCGGCGAGAGCATGCCATGGAT 4119
Qy 553 TCCAGTTTCGAGTATGGCGGCCAGAGGCTCCGACCTCGAGTGGCCATCAAAAGTCTTCTGCAACATGG 612
Db 4120 TCCAGTTTCGAGTATGGCGGCCAGAGGCTCCGACCTCGAGTGGCCATCAAAAGTCTTCTGCAACATGG 4179
```

QY	613	TCTTGGCGCTGATGTCTCACCGAGGCTCTCCAGAACATCACTTACCATCTGCAAGAACAGCG	672
Db	4180	TCCTTGGCGCTGATGTCTCACCGAGGCTCTCCAGAACATCACTTACCATCTGCAAGAACAGCG	4239
QY	673	TGCGCTACATGGACCGACGACACTTGGCAACCTCAAGAGGCCCTGCTCCTCAAGGGCTCCA	732
Db	4240	TGGCCTACATGGACCGACGACACTTGGCAACCTCAAGAGGCCCTGCTCCTCAAGGGCTCCA	4299
QY	733	ACGAGATCGAGATTCGCGCCGAGGGCAACAGCGCGCTTCACTACAGCGTCACTGTTCGATG	792
Db	4300	ACGAGATCGAGATTCGCGCCGAGGGCAACAGCGCGCTTCACTACAGCGTCACTGTTCGATG	4359
QY	793	GCTGCAAGACTCAACCGAGAGCTGGGGCAAGACAGTGATTGAATACAAACACCAAGT	852
Db	4360	GCTGCAAGACTCAACCGAGAGCTGGGGCAAGACAGTGATTGAATACAAACACCAAGT	4419
QY	853	CTTCCCGCCTGCCCATCATCATGTGGCCCCCTTGGACGTTGGTGCCTCCAGACCGAAT	912
Db	4420	CCTCCCGCCTGCCCATCATCATGTGGCCCCCTTGGACGTTGGTGCCTCCAGACCGAAT	4479
QY	913	TCGGCTTCGACGTTGGCGCCTGTCTGCTTCTGTAAACTCCCTCCATC	959
Db	4480	TCGGCTTCGACGTTGGCGCCTGTCTGCTTCTGTAAACTCCCTCCATC	4526
RESULT 14			
US-09-880-107-3946			
; Sequence 3946, Application US/09880107			
; Patent No. US20020142981A1			
; GENERAL INFORMATION:			
; APPLICANT: Horne, Darci T.			
; APPLICANT: Vockley, Joseph G.			
; APPLICANT: Scherf, Uwe			
; APPLICANT: Gene Logic, Inc.			
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer			
; FILE REFERENCE: 44921-5028-WO			
; CURRENT APPLICATION NUMBER: US/09/880,107			
; CURRENT FILING DATE: 2001-06-14			
; PRIOR APPLICATION NUMBER: US 60/211,379			
; PRIOR FILING DATE: 2000-06-14			
; PRIOR APPLICATION NUMBER: US 60/237,054			
; PRIOR FILING DATE: 2000-10-02			
; NUMBER OF SEQ ID NOS: 3950			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 3946			
; LENGTH: 6728			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
; FEATURE:			
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z74615			
US-09-880-107-3946			
Query Match 97.8%; Score 942.2; DB 3; Length 6728;			
Best Local Similarity 99.7%; Pred. No. 7.5e-255;			
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
QY	13	GATCTTAACGCTTCCCTGGCCCCCATTTGGGGCCCTCGTCTCTCGGGTCGCACCTGGTGATG	72
Db	3580	GATCAACGCTTCCCTGGGGCCCATTTGGGGCCCTCGTCTCTCGGGTCGCACCTGGTGATG	3639
QY	73	CTGGTCTCTGTTGGTCCCGCCGCGCTCTCTGGACCTCTCTGGTCCCGCTGGTCTCCACGCG	132
Db	3640	CTGGTCTCTGTTGGTCCCGCCGCGCTCTCTGGACCTCTCTGGTCCCGCTGGTCTCCACGCG	3699
QY	133	CTGGTTTCGACTTCAGCTTCTCTGCCCCCAGCCACTCAAGAGAGGCTCAAGATGGTGGCC	192
Db	3700	CTGGTTTCGACTTCAGCTTCTCTGCCCCCAGCCACTCAAGAGAGGCTCAAGATGGTGGCC	3759
QY	193	GCTTACTACCGGCTGATGATGCCAATGTGGTGTGTGACCGTGTGACCTCGAGGTGGAACCA	252
Db	3760	GCTTACTACCGGCTGATGATGCCAATGTGGTGTGTGACCGTGTGACCTCGAGGTGGAACCA	3819
QY	253	CCCTCAAGAGCCTGAGCCAGCAGATCGAAGACATCCCGAGCCAGAGGGGAAGCGCAAGA	312

Db	3820	CCCTCAAGAGCCTTGAGCCAGCAGATCGAAGCAATCGGAGCCAGAGGGAGAGTCGCAAGA	3879
Qy	313	ACCCGCGCCGACACTGCGGTGAGCTCTCAAGATGTGCCACTCTGTACTGGAAGAGTGGAGAGT	372
Db	3880	ACCCGCGCCGACACTGCGGTGAGCTCTCAAGATGTGCCACTCTGTACTGGAAGAGTGGAGAGT	3939
Qy	373	ACTGGATTGACCCCAACCAAGGCTGCAACCTTGGATGCCATCAAAGTCTTCTGCAACATGG	432
Db	3940	ACTGGATTGACCCCAACCAAGGCTGCAACCTTGGATGCCATCAAAGTCTTCTGCAACATGG	3999
Qy	433	AGACTGGTGGAGACTGGGTGTACCCACTCAGCCCAAGTGTGGGCCCAAGAACTGGTACA	492
Db	4000	AGACTGGTGGAGACTGGGTGTACCCACTCAGCCCAAGTGTGGGCCCAAGAACTGGTACA	4059
Qy	493	TCAGCAAGAACCCCAAGGCAAGAGGCATGTCTGGTTTCGGCGAGAGCATGACCCGATGGAT	552
Db	4060	TCAGCAAGAACCCCAAGGCAAGAGGCATGTCTGGTTTCGGCGAGAGCATGACCCGATGGAT	4119
Qy	553	TCCAGTTTCGAGTATGGCGGCGAGGGCTCCGACCTCTGCCGATGTGGCCATCCAGCTGACCT	612
Db	4120	TCCAGTTTCGAGTATGGCGGCGAGGGCTCCGACCTCTGCCGATGTGGCCATCCAGCTGACCT	4179
Qy	613	TCTTGGCGCTGATGTCTCACCGAGGCTCTCCAGAAATCATCACTTACACTGCAAGAAACAGCG	672
Db	4180	TCTTGGCGCTGATGTCTCACCGAGGCTCTCCAGAAATCATCACTTACACTGCAAGAAACAGCG	4239
Qy	673	TGCGCTTACATGGACACGACAGACTGGCAACTCAAGAAAGCCCTGTCTTCAAGGGGTCCA	732
Db	4240	TGCGCTTACATGGACACGACAGACTGGCAACTCAAGAAAGCCCTGTCTTCAAGGGGTCCA	4299
Qy	733	ACGAGATTCGAGATTCGCGCGCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGCGATG	792
Db	4300	ACGAGATTCGAGATTCGCGCGCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGCGATG	4359
Qy	793	GCTGCAAGAGTACACCGGAGCCTGGGGCAAGACAGTGAATGAATCAAAAACCAACAAGT	852
Db	4360	GCTGCAAGAGTACACCGGAGCCTGGGGCAAGACAGTGAATGAATCAAAAACCAACAAGT	4419
Qy	853	CCCTCCGCGCTGCCATCATGATGCGCCCTTGGACGCTTGGTGGCCCAAGACCAAGAT	912
Db	4420	CCCTCCGCGCTGCCATCATGATGCGCCCTTGGACGCTTGGTGGCCCAAGACCAAGAT	4479
Qy	913	TCGGCTTCGAGCTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC	959
Db	4480	TCGGCTTCGAGCTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC	4526
RESULT 15			
US-09-918-715-260			
; Sequence 260, Application US/09918715			
; Publication No. US20030017157A1			
; GENERAL INFORMATION:			
; APPLICANT: Brad St. Croix			
; APPLICANT: Bert Vogelstein			
; APPLICANT: Kenneth Kinzler			
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS			
; FILE REFERENCE: 1107.00134			
; CURRENT APPLICATION NUMBER: US/09/918,715			
; CURRENT FILING DATE: 2001-08-01			
; PRIOR APPLICATION NUMBER: 60/222,599			
; PRIOR FILING DATE: 2000-08-02			
; PRIOR APPLICATION NUMBER: 60/224,360			
; PRIOR FILING DATE: 2000-08-11			
; PRIOR APPLICATION NUMBER: 60/282,850			
; PRIOR FILING DATE: 2000-04-11			
; NUMBER OF SEQ ID NOS: 358			
; SOFTWARE: Fast-Seq for Windows Version 3.0			
; SEQ ID NO 260			
; LENGTH: 6728			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-918-715-260			

Query Match 97.8%; Score 942.2; DB 3; Length 6728;
Best Local Similarity 99.7%; Pred. No. 7.5e-255;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 13 GATCTAACGGTCTCCCTGGCCCAATTTGGCCCCCTGGTCTCTCGGGTGCACACTGGTGATG 72
Db |||||||
3580 GACTCAACGGTCTCCCTGGCCCAATTTGGCCCCCTGGTCTCTCGGGTGCACACTGGTGATG 3639

QY 73 CTGGTCTCTGTTGGTCCCGCCGCTCTCTGACCTCTCTGGTCCCGCTGGTCTCTCCAGCG 132
Db |||||||
3640 CTGGTCTCTGTTGGTCCCGCCGCTCTCTGACCTCTCTGGTCCCGCTGGTCTCTCCAGCG 3699

QY 133 CTGGTTCGACTTCAGCTTCTTGGCCCCCAGCCACCTCAAGAGAAGGCTCAAGATGTTGGCC 192
Db |||||||
3700 CTGGTTCGACTTCAGCTTCTTGGCCCCCAGCCACCTCAAGAGAAGGCTCAAGATGTTGGCC 3759

QY 193 GCTACTACCGGGCTGATGATGCAATGTGGTTCTGTGACCGTGAACCTCGAGGTGGACACCA 252
Db |||||||
3760 GCTACTACCGGGCTGATGATGCAATGTGGTTCTGTGACCGTGAACCTCGAGGTGGACACCA 3819

QY 253 CCCTCAAGAGCTGAGCCAGCAGATCGAGAAATCCGGAGCCCAAGAGGAAAGCCCAAGA 312
Db |||||||
3820 CCCTCAAGAGCTGAGCCAGCAGATCGAGAAATCCGGAGCCCAAGAGGAAAGCCCAAGA 3879

QY 313 ACCCGCCCGCACCTCGCGTGAACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGGT 372
Db |||||||
3880 ACCCGCCCGCACCTCGCGTGAACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGGT 3939

QY 373 ACTGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 432
Db |||||||
3940 ACTGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 3999

QY 433 AGACTGGTGAGACTGCGCTGTACCCCACTCAGCCAGTGTGGCCCAAGAACTGGTACA 492
Db |||||||
4000 AGACTGGTGAGACTGCGCTGTACCCCACTCAGCCAGTGTGGCCCAAGAACTGGTACA 4059

QY 493 TCAGCAAGAACCCCAAGGACAGAGGCGATGTCTGGTTGGCGAGAGCATGCCATGGAT 552
Db |||||||
4060 TCAGCAAGAACCCCAAGGACAGAGGCGATGTCTGGTTGGCGAGAGCATGCCATGGAT 4119

QY 553 TCCAGTTCCAGTATGCGCGCCAGAGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612
Db |||||||
4120 TCCAGTTCCAGTATGCGCGCCAGAGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 4179

QY 613 TCCTCGCGCTGATGTCCACCGAGGCTCCCAAGAACTCACCTACCACTGCAAGAACAGCG 672
Db |||||||
4180 TCCTCGCGCTGATGTCCACCGAGGCTCCCAAGAACTCACCTACCACTGCAAGAACAGCG 4239

QY 673 TGGCTTACATGGACAGCAGACTGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 732
Db |||||||
4240 TGGCTTACATGGACAGCAGACTGGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 4299

QY 733 ACGAGATCGAGATCGCGCCGAGGCGCAACGCCCTTCACTACAGGTCACTGTCGATG 792
Db |||||||
4300 ACGAGATCGAGATCGCGCCGAGGCGCAACGCCCTTCACTACAGGTCACTGTCGATG 4359

QY 793 GCTGCAAGGTACACCGGAGCTGGGGCAAGACAGATGATTGAATACAAACCAACCAAGT 852
Db |||||||
4360 GCTGCAAGGTACACCGGAGCTGGGGCAAGACAGATGATTGAATACAAACCAACCAAGT 4419

QY 853 CTTCCCGCTGCCCATCATCGATGTGGCCCTTGGAGTGGTGGTGGCCCGAGCCAGGAAT 912
Db |||||||
4420 CTTCCCGCTGCCCATCATCGATGTGGCCCTTGGAGTGGTGGTGGCCCGAGCCAGGAAT 4479

QY 913 TCGGCTTCGAGTTGGCCCTCTCTGCTTCTGTAATCTCCCTCCATC 959
Db |||||||
4480 TCGGCTTCGAGTTGGCCCTCTCTGCTTCTGTAATCTCCCTCCATC 4526

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 18:58:56 ; Search time 662.548 Seconds
(without alignments)
5902.968 Million cell updates/sec

Title: US-10-677-877A-1
Perfect score: 963
Sequence: 1 aagcttagtaagatctaac.....gtaaacctccatcatctaga 963

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:*

- 1: /SIDSS/ptodata/1/pubpna/US08 NEW PUB.seq.*
- 2: /SIDSS/ptodata/1/pubpna/US06 NEW PUB.seq.*
- 3: /SIDSS/ptodata/1/pubpna/US07 NEW PUB.seq.*
- 4: /SIDSS/ptodata/1/pubpna/PT NEW PUB.seq.*
- 5: /SIDSS/ptodata/1/pubpna/US09 NEW PUB.seq.*
- 6: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 7: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 8: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 9: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 10: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 11: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*
- 12: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*
- 13: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*
- 14: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*
- 15: /SIDSS/ptodata/1/pubpna/US60 NEW PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	942.2	97.8	5921	14	US-11-186-284-29
2	942.2	97.8	6728	9	US-10-501-035-138
3	942.2	97.8	6728	14	US-11-186-284-27
4	942.2	97.8	6728	14	US-11-021-603-1
5	937.4	97.3	4749	8	US-10-821-234-244
6	937.4	97.3	4770	14	US-11-000-463-7
7	937.4	97.3	4791	14	US-11-000-463-479
8	936	97.2	6741	14	US-11-091-883-105
9	936	97.2	6741	14	US-11-091-883-106
10	934.2	97.0	4752	7	US-10-784-004-1017
11	934.2	97.0	4752	7	US-10-784-004-1195
12	934.2	97.0	4752	7	US-10-784-004-81
13	801.4	83.2	5721	7	US-10-784-004-215
14	801.4	83.2	5721	7	US-10-784-004-826
15	801.4	83.2	5721	7	US-10-784-004-869
16	801.4	83.2	5721	7	US-10-784-004-869
17	786.4	81.7	5868	14	US-11-136-527-3354
18	654.2	67.9	659	14	US-11-108-172-991

19	636.2	66.1	642	14	US-11-108-172-966	Sequence 966, App
20	634.2	65.9	639	14	US-11-108-172-1003	Sequence 1003, App
21	632.2	65.6	637	14	US-11-108-172-1018	Sequence 1018, App
22	631.6	65.6	650	14	US-11-108-172-967	Sequence 967, App
23	631.2	65.5	636	14	US-11-108-172-1005	Sequence 1005, App
24	631.2	65.5	636	14	US-11-108-172-1022	Sequence 1022, App
25	623.2	64.7	629	14	US-11-108-172-1006	Sequence 1006, App
26	618.2	64.2	623	14	US-11-108-172-1019	Sequence 1019, App
27	616.2	64.0	621	14	US-11-108-172-1014	Sequence 1014, App
28	615.2	63.9	622	14	US-11-108-172-994	Sequence 994, App
29	612.2	63.6	629	14	US-11-108-172-968	Sequence 968, App
30	608.2	63.2	613	14	US-11-108-172-1002	Sequence 1002, App
31	593.8	61.7	609	14	US-11-108-172-972	Sequence 972, App
32	569.8	59.2	575	14	US-11-108-172-1007	Sequence 1007, App
33	563.8	58.5	576	14	US-11-108-172-920	Sequence 920, App
34	551.2	57.2	556	14	US-11-108-172-462	Sequence 462, App
35	490.2	50.9	5060	11	US-11-202-057-2	Sequence 2, Appli
36	490.2	50.9	5060	11	US-11-202-057-4	Sequence 4, Appli
37	490.2	50.9	5060	11	US-11-202-057-6	Sequence 6, Appli
38	481.6	50.0	4864	14	US-11-136-527-245	Sequence 245, App
39	454.4	47.2	461	14	US-11-108-172-985	Sequence 985, App
40	388.4	40.3	390	14	US-11-108-172-1001	Sequence 1001, App
41	348.8	36.2	5453	14	US-11-091-883-108	Sequence 108, App
42	347	36.0	5086	14	US-11-186-284-30	Sequence 30, Appl
43	347	36.0	5086	14	US-11-091-883-109	Sequence 109, App
44	342.8	35.6	4474	11	US-11-036-196-1986	Sequence 1986, App
45	340.6	35.4	4101	8	US-10-821-234-579	Sequence 579, App

ALIGNMENTS

RESULT 1

US-11-186-284-29
; Sequence 29, Application US/11/186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: HP01-029P2NM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR FILING DATE: 2002-11-21
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120) ... (4514)
US-11-186-284-29

Query Match 97.8%; Score 942.2; DB 14; Length 5921;
Best Local Similarity 99.7%; Pred. No. 2.8e-225; Indels 0; Gaps 0;
Matches 944; Conservative 0; Mismatches 3;


```
Db 4360 GCTGCAGAGTCAACCGGAGCTGGGCAAGACAGTGAATACAAACACCAAGT 4419
Qy 853 CTTCCCGCTGCCCATCATCATGATGGCCCTTGGAGCTGGTGGCCCGCCAGACAGGAAT 912
Db 4420 CTTCCCGCTGCCCATCATCATGATGGCCCTTGGAGCTGGTGGCCCGCCAGACAGGAAT 4479
Qy 913 TCGGCTTCGAGTTGGCCCTGCTGCTTCCTGTAACCTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTTGGCCCTGCTGCTTCCTGTAACCTCCCTCCATC 4526

RESULT 3
US-11-186-284-27
; Sequence 27, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPW01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)...(4514)
US-11-186-284-27

Query Match 97.8%; Score 942.2; DB 14; Length 6728;
Best Local Similarity 99.7%; Pred. No. 2.8e-225;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 13 GATCAACGGTCTCCCTGGCCCAATGGCCCTTGGCTCTCGGGTGGCACTGGTATG 72
Db 3580 GACTCAACGGTCTCCCTGGCCCAATGGCCCTTGGCTCTCGGGTGGCACTGGTATG 3639
Qy 73 CTGTGCTCTGTGGTCCCGCCGCTCTCTGACCTCTGTGGTCCCTGGTCTCCACAGCG 132
Db 3640 CTGTGCTCTGTGGTCCCGCCGCTCTCTGACCTCTGTGGTCCCTGGTCTCCACAGCG 3699
Qy 133 CTGTTTTCGACTTCAGCTTCTGTCGCCAGCCACTCAAGAGAAGGCTCAGATGGTGGCC 192
Db 3700 CTGTTTTCGACTTCAGCTTCTGTCGCCAGCCACTCAAGAGAAGGCTCAGATGGTGGCC 3759
Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGTTCTGACCGTGAACCTCAGGTGACACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCCAATGTGTTCTGACCGTGAACCTCAGGTGACACCA 3819
Qy 253 CCCTCAAGAGCTCAGGCCAGCAGATCGAGACATCCGGAGCCAGAGGAAGCCGCAAGA 312
Db 3820 CCCTCAAGAGCTCAGGCCAGCAGATCGAGACATCCGGAGCCAGAGGAAGCCGCAAGA 3879
```

```
Qy 313 ACCCGCCGCGACCTGCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAT 372
Db 3880 ACCCGCCGCGACCTGCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGG 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGG 3999
Qy 433 AGACTGCTGAGACTGCTGCTGTA CCCCACCTCAGCCAGTGTGGCCCAAGAACTGGGTACA 492
Db 4000 AGACTGCTGAGACTGCTGCTGTA CCCCACCTCAGCCAGTGTGGCCCAAGAACTGGGTACA 4059
Qy 493 TCAGCAAGAACCCCAAGGACAAGAGGCATCTCTGTTTCGGGCGAGAGCATGACCGATGAT 552
Db 4060 TCAGCAAGAACCCCAAGGACAAGAGGCATCTCTGTTTCGGGCGAGAGCATGACCGATGAT 4119
Qy 553 TCAGTTTCAGATATGGCGGCGCAGGGCTCCGACCTCGCCGATGTGGCCATCCAGCTGACCT 612
Db 4120 TCAGTTTCAGATATGGCGGCGCAGGGCTCCGACCTCGCCGATGTGGCCATCCAGCTGACCT 4179
Qy 613 TCTGCGCTGATGTCCACCGAGGCTTCCAGAACATCACTACCTGCAAGAACAGCG 672
Db 4180 TCTGCGCTGATGTCCACCGAGGCTTCCAGAACATCACTACCTGCAAGAACAGCG 4239
Qy 673 TGGCTCATGAGACCGAGCAGACTGGGCAACCTCAAGAGGCGCTCTCTCAAGGGCTCCA 732
Db 4240 TGGCTCATGAGACCGAGCAGACTGGGCAACCTCAAGAGGCGCTCTCTCAAGGGCTCCA 4299
Qy 733 ACCAGATTCAGATATCGGCGCGAGGCAACAGCGCTTCACTACAGGGTCACTGTCCATG 792
Db 4300 ACCAGATTCAGATATCGGCGCGAGGCAACAGCGCTTCACTACAGGGTCACTGTCCATG 4359
Qy 793 GCTGCACAGTCAACCGGAGCTTGGGCGAAGACAGTGAATTGAATACAAACACCAAGT 852
Db 4360 GCTGCACAGTCAACCGGAGCTTGGGCGAAGACAGTGAATTGAATACAAACACCAAGT 4419
Qy 853 CTTCGCGCTGCCCATCATCATGATGGCCCTTGGACCTTGGTGGCCCGAGACAGGAAT 912
Db 4420 CTTCGCGCTGCCCATCATCATGATGGCCCTTGGACCTTGGTGGCCCGAGACAGGAAT 4479
Qy 913 TCGGCTTCGAGCTTGGCGCTGCTGCTTCTCTGTAACCTCCCTCCATC 959
Db 4480 TCGGCTTCGAGCTTGGCGCTGCTGCTTCTCTGTAACCTCCCTCCATC 4526

RESULT 4
US-11-021-603-1
; Sequence 1, Application US/11021603
; Publication No. US20060003954A1
; GENERAL INFORMATION:
; APPLICANT: Berti, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/11/021,603
; CURRENT FILING DATE: 2004-12-21
; PRIOR APPLICATION NUMBER: US/10/149,352
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 1
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)..(4511)
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: (120)..(185)
```

US-11-021-603-1

Query Match 97.8%; Score 942.2; DB 14; Length 6728;
Best Local Similarity 99.7%; Pred. No. 2.8e-225;
Matches 944; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 13 GATCTAACGGTCTCCCTGGGCCCATTTGGGGCCCCCTGGTCTCGCGGTGCGACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGGCCCATTTGGGGCCCCCTGGTCTCGCGGTGCGACTGGTGATG 3639
Qy 73 CTGGTCTCTGTGGTCCCGCGGCCCTCTGAGACTCTGTGTCCTGTCCTGTCCTCCAGCG 132
Db 3640 CTGGTCTCTGTGGTCCCGCGGCCCTCTGAGACTCTGTGTCCTGTCCTCCAGCG 3699
Qy 133 CTGTTTTTCGACTTCAGCTTCTGCCCCCAGCCACTCAAGAGAGGCTCAGAGTGGGCC 192
Db 3700 CTGTTTTTCGACTTCAGCTTCTGCCCCCAGCCACTCAAGAGAGGCTCAGAGTGGGCC 3759
Qy 193 GCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 3819
Qy 253 CCCTCAAGAGCTGAGCCAGAGATCGAGAAATCGCGAGCCCGAGAGGCGGCAAGA 312
Db 3820 CCCTCAAGAGCTGAGCCAGAGATCGAGAAATCGCGAGCCCGAGAGGCGGCAAGA 3879
Qy 313 ACCCGCCCGACCTGCGGTGAGCTCAAGATGTGCCACTCTGACTGAGAGTGGAGAGT 372
Db 3880 ACCCGCCCGACCTGCGGTGAGCTCAAGATGTGCCACTCTGACTGAGAGTGGAGAGT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 3999
Qy 433 AGACTGGTGAAGTTCACCGAGGCTCCAGAAATCACTTCAAGAGTGGAGAGT 612
Db 4000 AGACTGGTGAAGTTCACCGAGGCTCCAGAAATCACTTCAAGAGTGGAGAGT 4059
Qy 493 TCAGCAGAGAACCCCAAGGAGCAAGGAGTCTGTGGTTCGGCGAGAGCATGACCGATGGAT 552
Db 4060 TCAGCAGAGAACCCCAAGGAGCAAGGAGTCTGTGGTTCGGCGAGAGCATGACCGATGGAT 4119
Qy 553 TCAGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGAGAGT 372
Db 4120 TCAGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGAGAGT 3939
Qy 613 TCCTGGCCCTGATGTCCACCGAGGCTCCAGAAATCACTTCAAGAGTGGAGAGT 672
Db 4180 TCCTGGCCCTGATGTCCACCGAGGCTCCAGAAATCACTTCAAGAGTGGAGAGT 4239
Qy 793 GCTGACAGTGTCAACCGAGGCTTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGT 852
Db 4360 GCTGACAGTGTCAACCGAGGCTTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGT 4419
Qy 853 CCTCCCGCTGCCCATCATGATGTGGCCCTTGGAGCTGGTGGCCCGAGACAGGAT 912
Db 4420 CCTCCCGCTGCCCATCATGATGTGGCCCTTGGAGCTGGTGGCCCGAGACAGGAT 4479
Qy 913 TCGGCTTCGAGTGGCCCTGTCTGTCTTCTGTAAATCTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTGGCCCTGTCTGTCTTCTGTAAATCTCCCTCCATC 4526

RESULT 5
US-10-821-234-244

; Sequence 244, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 244
; LENGTH: 4749
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-244

Query Match 97.3%; Score 937.4; DB 8; Length 4749;
Best Local Similarity 99.4%; Pred. No. 4.3e-224;
Matches 941; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 13 GATCTAACGGTCTCCCTGGGCCCATTTGGGGCCCCCTGGTCTCGCGGTGCGACTGGTGATG 72
Db 3580 GACTCAACGGTCTCCCTGGGCCCATTTGGGGCCCCCTGGTCTCGCGGTGCGACTGGTGATG 3639
Qy 73 CTGGTCTCTGTGGTCCCGCGGCCCTCTGAGACTCTGTGTCCTGTCCTCCAGCG 132
Db 3640 CTGGTCTCTGTGGTCCCGCGGCCCTCTGAGACTCTGTGTCCTGTCCTCCAGCG 3699
Qy 133 CTGGTTTTTCGACTTCAGCTTCTGCCCCCAGCCACTCAAGAGAGGCTCAGAGTGGGCC 192
Db 3700 CTGGTTTTTCGACTTCAGCTTCTGCCCCCAGCCACTCAAGAGAGGCTCAGAGTGGGCC 3759
Qy 193 GCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 252
Db 3760 GCTACTACCGGGCTGATGATGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 3819
Qy 253 CCCTCAAGAGCTGAGCCAGAGATCGAGAAATCGCGAGCCCGAGAGGCGGCAAGA 312
Db 3820 CCCTCAAGAGCTGAGCCAGAGATCGAGAAATCGCGAGCCCGAGAGGCGGCAAGA 3879
Qy 313 ACCCGCCCGACCTGCGGTGAGCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAGT 372
Db 3880 ACCCGCCCGACCTGCGGTGAGCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAGT 3939
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 432
Db 3940 ACTGGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 3999
Qy 433 AGACTGGTGAAGTTCACCGAGGCTCCAGAAATCACTTCAAGAGTGGAGAGT 612
Db 4000 AGACTGGTGAAGTTCACCGAGGCTCCAGAAATCACTTCAAGAGTGGAGAGT 4059
Qy 493 TCAGCAGAGAACCCCAAGGAGCAAGGAGTCTGTGGTTCGGCGAGAGCATGACCGATGGAT 552
Db 4060 TCAGCAGAGAACCCCAAGGAGCAAGGAGTCTGTGGTTCGGCGAGAGCATGACCGATGGAT 4119
Qy 553 TCAGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGAGAGT 612
Db 4120 TCAGATTGACCCCAACCAAGGCTGCAATGTGGTTCGTGACCGTGACCTCGAGGTGGAGAGT 4179
Qy 613 TCCTGGCCCTGATGTCCACCGAGGCTCCAGAAATCACTTCAAGAGTGGAGAGT 672
Db 4180 TCCTGGCCCTGATGTCCACCGAGGCTCCAGAAATCACTTCAAGAGTGGAGAGT 4239
Qy 793 GCTGACAGTGTCAACCGAGGCTTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGT 732
Db 4360 GCTGACAGTGTCAACCGAGGCTTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGT 4299
Qy 853 CCTCCCGCTGCCCATCATGATGTGGCCCTTGGAGCTGGTGGCCCGAGACAGGAT 912
Db 4420 CCTCCCGCTGCCCATCATGATGTGGCCCTTGGAGCTGGTGGCCCGAGACAGGAT 4479
Qy 913 TCGGCTTCGAGTGGCCCTGTCTGTCTTCTGTAAATCTCCCTCCATC 959
Db 4480 TCGGCTTCGAGTGGCCCTGTCTGTCTTCTGTAAATCTCCCTCCATC 4526

Qy	733	ACGAGATCGAGATCGGCGCCGAGGGCAACAGCCGCTTCACTCAAGCGTCACCTGTCGATG	792
Db	4300	ACGAGATCGAGATCGGCGCCGAGGGCAACAGCCGCTTCACTCAAGCGTCACCTGTCGATG	4359
Qy	793	GCTGCAAGAGTCACACCGAGGCTTGGGCGAAGACAGTATTGAATACAAACACCAACGAAT	852
Db	4360	GCTGCAAGAGTCACACCGAGGCTTGGGCGAAGACAGTATTGAATACAAACACCAACGAAT	4419
Qy	853	CCTCCCGCTGCCCCATCATCGATGTGGCCCCCTTTGGAGCTTTGGTGCCCCAGACACGGAAT	912
Db	4420	CCTCCCGCTGCCCCATCATCGATGTGGCCCCCTTTGGAGCTTTGGTGCCCCAGACACGGAAT	4479
Qy	913	TCGGCTTCGAGTTGGCCCTGTCTGCTTCGTGTAACCTCCCTCCATC	959
Db	4480	TCGGCTTCGAGTTGGCCCTGTCTGCTTCGTGTAACCTCCCTCCATC	4526

RESULT 6

```

US-11-000-463-7
; Sequence 7, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 4770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)..(4531)
US-11-000-463-7

```

Qy	133	CTGGTTTCGACTTCAGCTTCTCGCCCGACCACTCAAGAGAAGGCTCAGATGGGCC	192
Db	3720	CTGGTTTCGACTTCAGCTTCTCGCCCGACCACTCAAGAGAAGGCTCAGATGGTGCC	3779
Qy	193	GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGCTGCACCTCGAGTGGACACCA	252
Db	3780	GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGCTGCACCTCGAGTGGACACCA	3839
Qy	253	CCCTCAAGAGCTGAGCCGACGATCGAGAACAATTCGGGAGCCGACAGAGGGAAGCCGCAAGA	312
Db	3840	CCCTCAAGAGCTGAGCCGACGATCGAGAACAATTCGGGAGCCGACAGAGGGAAGCCGCAAGA	3899
Qy	313	ACCCGCGCCGCACTGCGCGTGACTCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT	372
Db	3900	ACCCGCGCGCACTGCGCGTGACTCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT	3959
Qy	373	ACTGGATTGACCCCAACCAAGGGTGCACACTGGATGCCATCAAAGTCTTCTGCAACAATGG	432
Db	3960	ACTGGATTGACCCCAACCAAGGGTGCACACTGGATGCCATCAAAGTCTTCTGCAACAATGG	4019
Qy	433	AGACTGTGTAGACCTTGGTGTATCCCACTCAGCCCGATGTGGCTGGCCCGAAGAACTGGTACA	492
Db	4020	AGACTGTGTAGACCTTGGTGTATCCCACTCAGCCCGATGTGGCTGGCCCGAAGAACTGGTACA	4079
Qy	493	TCAGCAAGAACCCTCAAGGACCAAGAGGCATGCTGTGTTTCGGCGGAGAGCATCACCGATGGAT	552
Db	4080	TCAGCAAGAACCCTCAAGGACCAAGAGGCATGCTGTGTTTCGGCGGAGAGCATCACCGATGGAT	4139
Qy	553	TCCAGTTTCGAGTATGGCGGCAGGGCTTCGACCTTCGCCGATGTGGCCATCCAGCTGACCT	612
Db	4140	TCCAGTTTCGAGTATGGCGGCAGGGCTTCGACCTTCGCCGATGTGGCCATCCAGCTGACCT	4199
Qy	613	TCCTGGCGCTGTATGTCACCGAGGCTCCGAGAACATCACCTTACCCTGCAAGAACAGCG	672
Db	4200	TCCTGGCGCTGTATGTCACCGAGGCTCCGAGAACATCACCTTACCCTGCAAGAACAGCG	4259
Qy	673	TGGCCTACATGAGACCAAGAGCTGGCAACCTCAAGAGGCGCTGCTCCTCAAGGGCTCCA	732
Db	4260	TGGCCTACATGAGACCAAGAGCTGGCACAACCTCAAGAGGCGCTGCTCCTCAAGGGCTCCA	4319
Qy	733	ACGAGATCGAGATCCGCGCCGAGGGCAACAGCGCTTACCTACAGGCTCACTGTGCGATG	792
Db	4320	ACGAGATCGAGATCCGCGCCGAGGGCAACAGCGCTTACCTACAGGCTCACTGTGCGATG	4379
Qy	793	GCTGCAAGAGTCAACCGGAGCTGGGGCAAGACAGTGAATGAATACAAACCAACCAAGT	852
Db	4380	GCTGCAAGAGTCAACCGGAGCTGGGGCAAGACAGTGAATGAATACAAACCAACCAAGT	4439
Qy	853	CCTCCGCCCTGCCCATCATCGATGTGGCCGCCCTTGGACGTTGGTGGCCCAAGACCAAGAA	912
Db	4440	CCTCCGCCCTGCCCATCATCGATGTGGCCGCCCTTGGACGTTGGTGGCCCAAGACCAAGAA	4499
Qy	913	TCGGCTTCGACGTTGGCCCTGTCTGCTTCTGTAAACTCCCTCATC	959
Db	4500	TCGGCTTCGACGTTGGCCCTGTCTGCTTCTGTAAACTCCCTCATC	4546

RESULT 7

US-011-000-463-479
; Sequence 479, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng

APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785C1P4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 479
; LENGTH: 4791
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(4791)
; OTHER INFORMATION: n = a,t,c or g
US-11-000-463-479

Query Match 97.3%; Score 937.4; DB 14; Length 4791;
Best Local Similarity 99.4%; Pred. No. 4.3e-224;
Matches 941; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 13 GATCTAACGGTCTCCCTGGCCCCCATGGGCCCTCTGGTCTCGCGTTCGCACTGGTGATG 72
Db 3620 GACTCAACGGTCTCCCTGGCCCCCATGGGCCCTCTGGTCTCGCGTTCGCACTGGTGATG 3679
Qy 73 CTGGTCTCTGTGGTCCCGCCCGCCCTCTGCACTCTCTGGTCCCTGGTCTCCCGAGCG 132
Db 3680 CTGGTCTCTGTGGTCCCGCCCGCCCTCTGCACTCTCTGGTCCCTGGTCTCCCGAGCG 3739
Qy 133 CTGGTTTCGACTTCAGCTTCTCTGCCCCCAGCCACTCAAGAGAGGCTCAGATGGTGGCC 192
Db 3740 CTGGTTTCGACTTCAGCTTCTCTGCCCCCAGCCACTCAAGAGAGGCTCAGATGGTGGCC 3799
Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCTGACCCGTGACCTCGAGTGGACACCA 252
Db 3800 GCTACTACCGGGCTGATGATGCCAATGTGGTTCTGACCCGTGACCTCGAGTGGACACCA 3859
Qy 253 CCCTCAAGAGCTGAGCCAGGATCGAGACATCCCGAGCCAGAGGAGCCGCAAGA 312
Db 3860 CCCTCAAGAGCTGAGCCAGGATCGAGACATCCCGAGCCAGAGGAGCCGCGCAAGA 3919
Qy 313 ACCCGCCGCACTCGCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAT 372
Db 3920 ACCCGCCGCACTCGCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAT 3979
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGTCATCAAGTCTTCTGCAACATGG 432
Db 3980 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGTCATCAAGTCTTCTGCAACATGG 4039
Qy 433 AGACTGTGAGACTGGTGTACCCCACTAGCCCACTGAGCCAGTGGCCCAAGACTGGTACA 492
Db 4040 AGACTGTGAGACTGGTGTACCCCACTAGCCCACTGAGCCAGTGGCCCAAGACTGGTACA 4099
Qy 493 TCAGCAAGAGCCCAAGAGGATGTCTGGTTCGGGAGAGATGACCATGATGGAT 552
Db 4100 TCAGCAAGAGCCCAAGAGGATGTCTGGTTCGGGAGAGATGATGATGATGAT 4159
Qy 553 TCAGATTGAGATGGCGGCGAGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612
Db 553 TCAGATTGAGATGGCGGCGAGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612

RESULT 8
US-11-091-883-105/c
; Sequence 105, Application US/11091883
; Publication No. US20060024693A1
; GENERAL INFORMATION:
; APPLICANT: CIBELLI, JOSE
; APPLICANT: FERNANDEZ, EMILIO O.
; APPLICANT: JORDAO DE MEGALHAES, GUILHERME
; APPLICANT: KOCABAS, ARIF
; APPLICANT: CROSBY, JAVIER A.
; TITLE OF INVENTION: IDENTIFICATION OF GENES OR POLYPEPTIDES THE EXPRESSION OF WHICH
; TITLE OF INVENTION: CORRELATES TO FERTILITY, OVARIAN FUNCTION AND/OR FETAL/NEBORN
; TITLE OF INVENTION: VIABILITY
; FILE REFERENCE: S3942US
; CURRENT APPLICATION NUMBER: US/11/091,883
; CURRENT FILING DATE: 2005-03-29
; PRIOR APPLICATION NUMBER: 60/556,875
; PRIOR FILING DATE: 2004-03-29
; NUMBER OF SEQ ID NOS: 513
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
; LENGTH: 6741
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (758)..(758)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (760)..(760)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (762)..(766)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (768)..(775)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (777)..(782)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base

Db 4160 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCCCTGCCGATGTGGCCATCCAGCTGACCT 4219
Qy 613 TCCTGCGCCTGATGTCCACCGAGGCTTCCAGAGAACTCCTTACCTACCTGCAAGAACAGCG 672
Db 4220 TCCTGCGCCTGATGTCCACCGAGGCTTCCAGAGAACTCCTTACCTACCTGCAAGAACAGCG 4279
Qy 673 TGGCCTACATGGACACGAGACTGGCAACCTCAAGAGGCTTCTGCTCTCAAGGGCTTCCA 732
Db 4280 TGGCCTACATGGACACGAGACTGGCAACCTCAAGAGGCTTCTGCTCTCAAGGGCTTCCA 4339
Qy 733 ACAGATCGAGATCCCGCGGAGGCAACAGCCGCTTACCTACAGGCTCACTGTGATG 792
Db 4340 ACAGATCGAGATCCCGCGGAGGCAACAGCCGCTTACCTACAGGCTCACTGTGATG 4399
Qy 793 GCTGCACGAGTCAACCGGAGCTGGGGCAAGACAGTGTGAATACAAACACCAAGT 852
Db 4400 GCTGCACGAGTCAACCGGAGCTGGGGCAAGACAGTGTGAATACAAACACCAAGT 4459
Qy 853 CTCTCCCGCTGCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCCAGACAGGAAT 912
Db 4460 CTCTCCCGCTGCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCCAGACAGGAAT 4519
Qy 913 TCGGCTTCGACGTTGGCCCCCTGCTTCTGTAAAACCTCCCTCCATC 959
Db 4520 TCGGCTTCGACGTTGGCCCCCTGCTTCTGTAAAACCTCCCTCCATC 4566

```
LOCATION: (810)..(810)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (812)..(812)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (928)..(928)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1203)..(1203)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1207)..(1207)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1420)..(1424)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1485)..(1485)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1501)..(1501)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1542)..(1545)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1692)..(1692)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1841)..(1841)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1892)..(1892)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1931)..(1938)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (1940)..(1948)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (2009)..(2009)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (2871)..(2871)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (2873)..(2873)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (3020)..(3020)
OTHER INFORMATION: a, c, g, or t
FEATURE:
NAME/KEY: modified base
LOCATION: (3380)..(3380)
```

```
OTHER INFORMATION: a, c, g, or t
us-11-091-883-105
```

```
Query Match 97.2%; Score 936; DB 14; Length 6741;
Best Local Similarity 99.2%; Pred. No. 1e-223;
Matches 939; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 13 GATCTAACGGTCTCCCTGGGCCCATTTGGGCCCATTTGGGCCCATTTGGTCTCTCGGGTTCGCACTGGTCAATG 72
Db 3162 GACTCAACGGTCTCCCTGGGCCCATTTGGGCCCATTTGGTCTCTCGGGTTCGCACTGGTCAATG 3103

Qy 73 CTGGTCTCTGTTGGTTCCTCGGCCCTCTCTGACCTCTCTGGTTCCTCGCTCTCTCGCTCTCTCGAGCG 132
Db 3102 CTGGTCTCTGTTGGTTCCTCGGCCCTCTCTGACCTCTCTGGTTCCTCGCTCTCTCGAGCG 3043

Qy 133 CTGGTTTCGACTTCAGCTTCTCTGCCCGCCCACTCAAGAGAGAGGCTCACGATGGTGGCC 192
Db 3042 CTGGTTTCGACTTCAGCTTCTCTGCCCGCCCACTCAAGAGAGAGGCTCACGATGGTGGCC 2983

Qy 193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGCCTCGAGGTGACACCA 252
Db 2982 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGCCTCGAGGTGACACCA 2923

Qy 253 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATTCGGAGCCCGAGAGGGAAGCGCAAGA 312
Db 2922 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATTCGGAGCCCGAGAGGGAAGCGCAAGA 2863

Qy 313 ACCCGCCCGCACCTTGCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 372
Db 2862 ACCCGCCCGCACCTTGCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 2803

Qy 373 ACTGGATTGACCCCAACCAAGGCTGCAACTGTGATGCCATCAAAAGTCTTCTGCAACATGG 432
Db 2802 ACTGGATTGACCCCAACCAAGGCTGCAACTGTGATGCCATCAAAAGTCTTCTGCAACATGG 2743

Qy 433 AGACTGTGAGACTTCGGTGTACCCCACTCAGAGCCAGTGTGGCCCGAGAGAACTGTGTACA 492
Db 2742 AGACTGTGAGACTTCGGTGTACCCCACTCAGAGCCAGTGTGGCCCGAGAGAACTGTGTACA 2693

Qy 493 TCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGGTTTCGGCGAGAGCATCACCGATGGAT 552
Db 2682 TCAGCAAGAACCCCAAGGCAAGAGGCAATGTCTGGTTTCGGCGAGAGCATCACCGATGGAT 2623

Qy 553 TCAGTTTCGAGTATGGCGGCGCAGGGCTTCGACCTCGCCGATGTGGCCATCCAGTGAACCT 612
Db 2622 TCAGTTTCGAGTATGGCGGCGCAGGGCTTCGACCTCGCCGATGTGGCCATCCAGTGAACCT 2563

Qy 613 TCCTGGCCCTGATGTCCACCGAGGCTCCAGAGCAATCACCTACCACTGCAAGAACAGCG 672
Db 2562 TCCTGGCCCTGATGTCCACCGAGGCTCCAGAGCAATCACCTACCACTGCAAGAACAGCG 2503

Qy 673 TGGCCTACATGGACCAAGAGACTGGCAACCTCAAGAGGCGCTCTGCTCTCAAGGGCTTCCA 732
Db 2502 TGGCCTACATGGACCAAGAGACTGGCAACCTCAAGAGGCGCTCTGCTCTCAAGGGCTTCCA 2443

Qy 733 ACAGATTCGAGATTCGGCGGCGAGGGCAACAGCCGCTTCACTACAGGTCACTGTGATG 792
Db 2442 ACAGATTCGAGATTCGGCGGCGAGGGCAACAGCCGCTTCACTACAGGTCACTGTGATG 2383

Qy 793 GCTGCACGAGTTCACCGGAGCCTGGGCGAAGCAGTGTGATTAACAAACCAACCAAGT 852
Db 2382 GCTGCACGAGTTCACCGGAGCCTGGGCGAAGCAGTGTGATTAACAAACCAACCAAGT 2323

Qy 853 CCTCCCGCCTGCCCATCATCGATGTGGCCCTTGGACGTTGGTGGCCCGAGACAGGAAT 912
Db 2322 CCTCCCGCCTGCCCATCATCGATGTGGCCCTTGGACGTTGGTGGCCCGAGACAGGAAT 2263

Qy 913 TCGGCTTCGAGCTTGGCCCTGTCTGTCTCTGTAAACTCCCTCCATC 959
Db 2262 TCGGCTTCGAGCTTGGCCCTGTCTGTCTCTGTAAACTCCCTCCATC 2216
```

RESULT 9

US-11-091-883-106/c
; Sequence 106, Application US/11091883
; Publication No. US20060024693A1
; GENERAL INFORMATION:
; APPLICANT: CIBELLI, JOSE
; APPLICANT: FERNANDEZ, EMILIO O.
; APPLICANT: JORDAO DE MEGALHAES, GUILHERME
; APPLICANT: KOCABAS, ARIFF
; APPLICANT: CROSBY, JAVIER A.
; TITLE OF INVENTION: IDENTIFICATION OF GENES OR POLYPEPTIDES THE EXPRESSION OF WHICH
; TITLE OF INVENTION: CORRELATES TO FERTILITY, OVARIAN FUNCTION AND/OR FETAL/NEBORN
; TITLE OF INVENTION: VIABILITY
; FILE REFERENCE: 53942US
; CURRENT APPLICATION NUMBER: US/11/091,883
; CURRENT FILING DATE: 2005-03-29
; PRIOR APPLICATION NUMBER: 60/556,875
; PRIOR FILING DATE: 2004-03-29
; NUMBER OF SEQ ID NOS: 513
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 106
; LENGTH: 6741
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (758)..(758)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (760)..(760)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (762)..(766)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (768)..(775)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (777)..(782)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (810)..(810)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (812)..(812)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (928)..(928)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1203)..(1203)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1207)..(1207)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1420)..(1424)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1485)..(1485)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base

; LOCATION: (1501)..(1501)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1542)..(1545)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1692)..(1692)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1841)..(1841)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1892)..(1892)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1931)..(1938)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1940)..(1948)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2009)..(2009)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2871)..(2871)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2873)..(2873)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (3020)..(3020)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (3380)..(3380)
; OTHER INFORMATION: a, c, g, or t
; US-11-091-883-106

Query Match 97.2%; Score 936; DB 14; Length 6741;
Best Local Similarity 99.2%; Pred. No. 1e-223;
Matches 939; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 13 GATCTAACGGTCTCCCTGGCCCCCATTTGGGGCCCCCTCGTCTCGCGGTCCGACCTGGTGATG 72
Db 3162 GACTCAACGGTCTCCCTGGCCCCCATTTGGGGCCCCCTCGTCTCGCGGTCCGACCTGGTGATG 3103
QY 73 CTGGTCTCTGTTGGTCTCCCGCGGCCCTCTGGACCTCTCTGGTCTCCCTGGTCTCCAGCG 132
Db 3102 CTGGTCTCTGTTGGTCTCCCGCGGCCCTCTGGACCTCTCTGGTCTCCCTGGTCTCCAGCG 3043
QY 133 CTGGTTTCGACCTTCAGCTTCTCTGCCCCCAGCCACCTCAAGAGAGGGCTCAGATGGTGGCC 192
Db 3042 CTGGTTTCGACCTTCAGCTTCTCTGCCCCCAGCCACCTCAAGAGAGGGCTCAGATGGTGGCC 2983
QY 193 GCTACTACCGGGCTGATGATGCCAAATGTGGTTGGTACCGGTGACCTCGAGGTGCACCA 252
Db 2982 GCTACTACCGGGCTGATGATGCCAAATGTGGTTGGTACCGGTGACCTCGAGGTGCACCA 2923
QY 253 CCCTCAAGAGCCCTGAGCCAGCAGATCGAGAAATCCGGAGCCCGAGAGGGAAGCCGCAAGA 312
Db 2922 CCCTCAAGAGCCCTGAGCCAGCAGATCGAGAAATCCGGAGCCCGAGAGGGAAGCCGCAAGA 2863
QY 313 ACCCCGCCCCGACCTCTCCCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 372

Db 2862 ACCCGCCGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT 2803
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 432
Db 2802 ACTGGATTGACCCCAACCAAGGCTGCACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 2743
Qy 433 AGACTGTGTGACACTCGGTGTACCCACTCAGCCAGTGTGCCAGAGAACTGGTACA 492
Db 2742 AGACTGTGTGACACTCGGTGTACCCACTCAGCCAGTGTGCCAGAGAACTGGTACA 2693
Qy 493 TCAGCAAGAACCCCAAGCAAGAGGATGTCTGTTTCGGCGAGAGATGACCCATGGAT 552
Db 2682 TCAGCAAGAACCCCAAGCAAGAGGATGTCTGTTTCGGCGAGAGATGACCCATGGAT 2623
Qy 553 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612
Db 2622 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 2563
Qy 613 TCCTGGCGCTGATGTCAACGAGGCTCCAGAAATCATCCTACACTGCAAGAAACAGCG 672
Db 2562 TCCTGGCGCTGATGTCAACGAGGCTCCAGAAATCATCCTACACTGCAAGAAACAGCG 2503
Qy 673 TGGCCTACATGGACGACGACTGGCAACTCCTCAGAGAGGCTCTCTCAAGGGCTCCA 732
Db 2502 TGGCCTACATGGACGACGACTGGCAACTCCTCAGAGAGGCTCTCTCTCAAGGGCTCCA 2443
Qy 733 ACAGAGATCGAGATCCGGGCCAGGGCAACAGCCGCTTCACTACAGAGTCACTGTCCATG 792
Db 2442 ACAGAGATCGAGATCCGGGCCAGGGCAACAGCCGCTTCACTACAGAGTCACTGTCCATG 2393
Qy 793 GCTGCACGAGTCAACCGGAGCTTGGGGCAAGACAGTGAATTAATAAAAAACCAACCAAGT 852
Db 2382 GCTGCACGAGTCAACCGGAGCTTGGGGCAAGACAGTGAATTAATAAAAAACCAACCAAGT 2323
Qy 853 CCTCCGCTGCCATCATCATGATGTGGCCCTTGGAGCTTGGGCGGAGGAGGAGGAGGAGT 912
Db 2322 CCTCCGCTGCCATCATCATGATGTGGCCCTTGGAGCTTGGGCGGAGGAGGAGGAGGAGT 2263
Qy 913 TCGGCTTCGAGTGTGGCCCTCTCTGCTTCTCTGTAATCTCCCTCCATC 959
Db 2262 TCGGCTTCGAGTGTGGCCCTCTCTGCTTCTCTGTAATCTCCCTCCATC 2216

RESULT 10

US-10-784-004-583
; Sequence 583, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:

; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 583

; LENGTH: 4752

; TYPE: DNA

; ORGANISM: human

US-10-784-004-583

Query Match 97.0%; Score 934.2; DB 7; Length 4752;

Best Local Similarity 99.2%; Pred. No. 2.7e-223;

Matches 939; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 13 GATCTAACGGTCTCCCTGGGCCCATATGGGCCCTTGGTCTCTCGGGTGGCACTGGTATG 72
Db 3577 GACTCAACGGTCTCCCTGGGCCCATATGGGCCCTTGGTCTCTCGGGTGGCACTGGTATG 3636

Qy 73 CTGGTCTCTGTGGTGGGCCCTCTCTGACCTCTGGTCCCTGGTCTCTCCAGCG 132
Db 3637 CTGGTCTCTGTGGTGGGCCCTCTCTGACCTCTGGTCTCTCTCCAGCG 3696

Qy 133 CTGGTTCGACTTCAGCTTCTCTGCCCCAGGCCACTCTCAAGAGAAAGGCTCACGATGGTGCC 192
Db 3697 CTGGTTCGACTTCAGCTTCTCTGCCCCAGGCCACTCTCAAGAGAAAGGCTCACGATGGTGCC 3756
Qy 193 GCTACTACCGGGCTGATGTCGCAATGTGTTCTGTGACCGTGTGACCTCGAGGTGGACACCA 252
Db 3757 GCTACTACCGGGCTGATGTCGCAATGTGTTCTGTGACCGTGTGACCTCGAGGTGGACACCA 3816
Qy 253 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATTCGGAGCCCAAGAGGAAAGCCGCAAGA 312
Db 3817 CCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATTCGGAGCCCAAGAGGAGCCGCAAGA 3876
Qy 313 ACCCGCCCGGACCTGCGGTGA CCTCAAGATGTGCCA CTCTGACTGGAAGAGTGGAGAGT 372
Db 3877 ACCCGCCCGGACCTGCGGTGA CCTCAAGATGTGCCA CTCTGACTGGAAGAGTGGAGAGT 3936
Qy 373 ACTGGATTGACCCCAACCAAGGCTGCACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 432
Db 3937 ACTGGATTGACCCCAACCAAGGCTGCACCTGGATGCCATCAAAAGTCTTCTGCAACATGG 3996
Qy 433 AGACTGTGTGAGACTGTGCTGTGACCCCACTCAGCCAGTGTGGCCCAAGAACTGGTACA 492
Db 3997 AGACTGTGTGAGACTGTGCTGTGACCCCACTCAGCCAGTGTGGCCCAAGAACTGGTACA 4056
Qy 493 TCAGCAAGAACCCCAAGCAAGAGGATGTCTGTTTCGGCGAGAGATGACCCATGGAT 552
Db 4057 TCAGCAAGAACCCCAAGCAAGAGGATGTCTGTTTCGGCGAGAGATGACCCATGGAT 4116
Qy 553 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 612
Db 4117 TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCT 4176
Qy 613 TCCTGCGCTGATGTCCACCGAGGCTTCCCAAGAACATCACCTACACTGCAAGAAACAGCG 672
Db 4177 TCCTGCGCTGATGTCCACCGAGGCTTCCCAAGAACATCACCTACACTGCAAGAAACAGCG 4236
Qy 673 TGGCCTACATGGACGACGACTGGGCAACTCTCAAGAGGCGCTCTCTCAAGGGCTCCA 732
Db 4237 TGGCCTACATGGACGACGACTGGGCAACTCTCAAGAGGCGCTCTCTCTCAAGGGCTCCA 4296
Qy 733 ACAGAGATCGAGATCCGGGCCAGGGCAACAGCCGCTTCACTACAGGCTCACTGTCCATG 792
Db 4297 ACAGAGATCGAGATCCGGGCCAGGGCAACAGCCGCTTCACTACAGGCTCACTGTCCATG 4356
Qy 793 GCTGCACGAGTCAACCGGAGCCTTGGGGCAAGACAGTGAATTAATAAAAAACCAACCAAGT 852
Db 4357 GCTGCACGAGTCAACCGGAGCCTTGGGGCAAGACAGTGAATTAATAAAAAACCAACCAAGT 4416
Qy 853 CCTCCGCTGCCATCATCATGATGTGGCCCTTGGAGCTTGGGCGGAGGAGGAGGAGGAGT 912
Db 4417 CCTCCGCTGCCATCATCATGATGTGGCCCTTGGAGCTTGGGCGGAGGAGGAGGAGGAGT 4476
Qy 913 TCGGCTTCGAGCTTGGCCCTCTCTGCTTCTCTGTAATCTCCCTCCATC 959
Db 4477 TCGGCTTCGAGCTTGGCCCTCTCTGCTTCTCTGTAATCTCCCTCCATC 4523

RESULT 11

US-10-784-004-1017

; Sequence 1017, Application US/10784004

; Publication No. US20060084066A1

; GENERAL INFORMATION:

; APPLICANT: Biogen Idec

; TITLE OF INVENTION: Surrogate Markers of Pain

; FILE REFERENCE: 08201.6029-00000

; CURRENT APPLICATION NUMBER: US/10/784,004

; CURRENT FILING DATE: 2004-02-20

; NUMBER OF SEQ ID NOS: 1251

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 1017

; LENGTH: 4752

; TYPE: DNA

US-10-784-004-1195

Query Match 97.0%; Score 934.2; DB 7; Length 4752;
Best Local Similarity 99.2%; Pred. No. 2.7e-223;
Matches 939; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

13 GATCTAACGGTCTCCCTGGGCCCAATTGGGGCCCTCGTCTCGCGTCCGACTGGTGATG 72
3577 GACTCAACGGTCTCCCTGGGCCCAATTGGGGCCCTCGTCTCGCGTCCGACTGGTGATG 3636

73 CTGGTCTCTGTGGTCCCGCCCGCCCTCTCGAGACTCTCTGGTCCCGCTGCTCTCCAGCG 132
3637 CTGGTCTCTGTGGTCCCGCCCGCCCTCTCGAGACTCTCTGGTCCCGCTGCTCTCCAGCG 3696

133 CTGGTTTCGACTTCAGCTTCTGCGCCCGCCCACTCAAGAGAAGGCTCAGATGGTGGCC 192
3697 CTGGTTTCGACTTCAGCTTCTGCGCCCGCCCACTCAAGAGAAGGCTCAGATGGTGGCC 3756

193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 252
3757 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 3816

253 CCCTCAAGAGCCTTGAGCAGAGATCGAGAACATCCGGAGCCCGAGAGGAGCGCAAGA 312
3817 CCCTCAAGAGCCTTGAGCAGAGATCGAGAACATCCGGAGCCCGAGAGGAGCGCGAAGA 3876

313 ACCCGCCCGCAGCTGCGCTGACCTCAAGATGTCACCTCAAGAGAAGGCTCAGATGGTGGCC 372
3877 ACCCGCCCGCAGCTGCGCTGACCTCAAGATGTCACCTCAAGAGAAGGCTCAGATGGTGGCC 3936

193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 252
3757 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 3816

253 CCCTCAAGAGCCTTGAGCAGAGATCGAGAACATCCGGAGCCCGAGAGGAGCGCAAGA 312
3817 CCCTCAAGAGCCTTGAGCAGAGATCGAGAACATCCGGAGCCCGAGAGGAGCGCGAAGA 3876

313 ACCCGCCCGCAGCTGCGCTGACCTCAAGATGTCACCTCAAGAGAAGGCTCAGATGGTGGCC 372
3877 ACCCGCCCGCAGCTGCGCTGACCTCAAGATGTCACCTCAAGAGAAGGCTCAGATGGTGGCC 3936

373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGG 432
3937 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGG 4416

433 AGACTGTGAGAGCTGCTGCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACCA 492
3997 AGACTGTGAGAGCTGCTGCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACCA 4056

493 TCAGCAAGAACCCCAAGGACCAAGAGGATGTCGTGGTTCGGCGAGAGATGACCGATGGAT 552
4057 TCAGCAAGAACCCCAAGGACCAAGAGGATGTCGTGGTTCGGCGAGAGATGACCGATGGAT 4116

553 TCCAGTTTCGAGTATGGCGCCGAGGCTCCGACCTCGCGATGGCCATCCAGTCTGACCTGACCT 612
4117 TCCAGTTTCGAGTATGGCGCCGAGGCTCCGACCTCGCGATGGCCATCCAGTCTGACCTGACCT 4176

433 AGACTGTGAGAGCTGCTGCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACCA 492
3997 AGACTGTGAGAGCTGCTGCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACCA 4056

493 TCAGCAAGAACCCCAAGGACCAAGAGGATGTCGTGGTTCGGCGAGAGATGACCGATGGAT 552
4057 TCAGCAAGAACCCCAAGGACCAAGAGGATGTCGTGGTTCGGCGAGAGATGACCGATGGAT 4116

553 TCCAGTTTCGAGTATGGCGCCGAGGCTCCGACCTCGCGATGGCCATCCAGTCTGACCTGACCT 612
4117 TCCAGTTTCGAGTATGGCGCCGAGGCTCCGACCTCGCGATGGCCATCCAGTCTGACCTGACCT 4176

613 TCCGCGCTGATGTCACCGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCG 672
4177 TCCGCGCTGATGTCACCGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCG 4236

73 TGGCTTACATGGACAGCAGATGCGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 732
4237 TGGCTTACATGGACAGCAGATGCGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 4296

733 ACCAGATTCGAGATCCCGCGGAGGCAACCGCTTCACTACAGGCTCACTGTGATG 792
4297 ACCAGATTCGAGATCCCGCGGAGGCAACCGCTTCACTACAGGCTCACTGTGATG 4356

793 GCTGCGACGATCAACCGGAGCCTGGGGCAAGACAGTGAATTGAATACAAACCAACCAAGT 852
4357 GCTGCGACGATCAACCGGAGCCTGGGGCAAGACAGTGAATTGAATACAAACCAACCAAGT 4416

853 CTCCCGGCTGCCCATCATGATGTGGCCCGCTTGGAGCTGGTGGCCCGAGACAGGAAT 912
4417 CTCCCGGCTGCCCATCATGATGTGGCCCGCTTGGAGCTGGTGGCCCGAGACAGGAAT 4476

913 TCGGCTTCGAGCTTGGCCCTGCTGCTTCTGTAACCTCCCTCCATC 959
4477 TCGGCTTCGAGCTTGGCCCTGCTGCTTCTGTAACCTCCCTCCATC 4523

US-10-784-004-1195

Sequence 1195, Application US/10784004
Publication No. US20060084066A1
GENERAL INFORMATION:
APPLICANT: Biogen Idec
TITLE OF INVENTION: Surrogate Markers of Pain
FILE REFERENCE: 08201.6029-00000
CURRENT APPLICATION NUMBER: US/10/784,004
CURRENT FILING DATE: 2004-02-20
NUMBER OF SEQ ID NOS: 1251
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1195
LENGTH: 4752
TYPE: DNA
ORGANISM: human
US-10-784-004-1195

Query Match 97.0%; Score 934.2; DB 7; Length 4752;
Best Local Similarity 99.2%; Pred. No. 2.7e-223;
Matches 939; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

13 GATCTAACGGTCTCCCTGGGCCCAATTGGGGCCCTCGTCTCGCGTCCGACTGGTGATG 72
3577 GACTCAACGGTCTCCCTGGGCCCAATTGGGGCCCTCGTCTCGCGTCCGACTGGTGATG 3636

73 CTGGTCTCTGTGGTCCCGCCCGCCCTCTCGAGACTCTCTGGTCCCGCTGCTCTCCAGCG 132
3637 CTGGTCTCTGTGGTCCCGCCCGCCCTCTCGAGACTCTCTGGTCCCGCTGCTCTCCAGCG 3696

133 CTGGTTTCGACTTCAGCTTCTGCGCCCGCCCACTCAAGAGAAGGCTCAGATGGTGGCC 192
3697 CTGGTTTCGACTTCAGCTTCTGCGCCCGCCCACTCAAGAGAAGGCTCAGATGGTGGCC 3756

193 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 252
3757 GCTACTACCGGGCTGATGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCA 3816

253 CCCTCAAGAGCCTTGAGCAGAGATCGAGAACATCCGGAGCCCGAGAGGAGCGCAAGA 312
3817 CCCTCAAGAGCCTTGAGCAGAGATCGAGAACATCCGGAGCCCGAGAGGAGCGCGAAGA 3876

313 ACCCGCCCGCAGCTGCGCTGACCTCAAGATGTCACCTCAAGAGAAGGCTCAGATGGTGGCC 372
3877 ACCCGCCCGCAGCTGCGCTGACCTCAAGATGTCACCTCAAGAGAAGGCTCAGATGGTGGCC 3936

373 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGG 432
3937 ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGG 3996

433 AGACTGTGAGAGCTGCTGCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACCA 492
3997 AGACTGTGAGAGCTGCTGCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACCA 4056

493 TCAGCAAGAACCCCAAGGACCAAGAGGATGTCGTGGTTCGGCGAGAGATGACCGATGGAT 552
4057 TCAGCAAGAACCCCAAGGACCAAGAGGATGTCGTGGTTCGGCGAGAGATGACCGATGGAT 4116

553 TCCAGTTTCGAGTATGGCGCCGAGGCTCCGACCTCGCGATGGCCATCCAGTCTGACCTGACCT 612
4117 TCCAGTTTCGAGTATGGCGCCGAGGCTCCGACCTCGCGATGGCCATCCAGTCTGACCTGACCT 4176

613 TCCGCGCTGATGTCACCGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCG 672
4177 TCCGCGCTGATGTCACCGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCG 4236

73 TGGCTTACATGGACAGCAGATGCGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 732
4237 TGGCTTACATGGACAGCAGATGCGCAACCTCAAGAGGCGCTGCTCTCAAGGGCTCCA 4296

733 ACCAGATTCGAGATCCCGCGGAGGCAACCGCTTCACTACAGGCTCACTGTGATG 792
4297 ACCAGATTCGAGATCCCGCGGAGGCAACCGCTTCACTACAGGCTCACTGTGATG 4356

793 GCTGCGACGATCAACCGGAGCCTGGGGCAAGACAGTGAATTGAATACAAACCAACCAAGT 852
4357 GCTGCGACGATCAACCGGAGCCTGGGGCAAGACAGTGAATTGAATACAAACCAACCAAGT 4416

853 CTCCCGGCTGCCCATCATGATGTGGCCCGCTTGGAGCTGGTGGCCCGAGACAGGAAT 912
4417 CTCCCGGCTGCCCATCATGATGTGGCCCGCTTGGAGCTGGTGGCCCGAGACAGGAAT 4476

913 TCGGCTTCGAGCTTGGCCCTGCTGCTTCTGTAACCTCCCTCCATC 959
4477 TCGGCTTCGAGCTTGGCCCTGCTGCTTCTGTAACCTCCCTCCATC 4523

QY	373	ACTGGATTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGG	432
Db	3788	ACTGGATCGAGCCCTAAACCAAGGCTGCAACCTGGATGCCATCAAGGCTTACTGCAACATGG	3847
QY	433	AGACTGGTGAACCTGGGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGGTACA	492
Db	3848	AGACAGGTTCAGACTGTGTGTTCCTCCCACTCAGCCCTGTGTGCTCAGAGAACTGGGTACA	3907
QY	493	TCAGCAAGAACCCCAAGGCAACAGAGGATGCTGTGGTTTCGGCGAGAGATGACCCGATGGAT	552
Db	3908	TCAGCCCAACCCCAAGGAGAAAGCATGTCTGGTTTGGAGAGAGATGACCCGATGGAT	3967
QY	553	TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTGCGCATGTGGCCATCCAGCTGACCT	612
Db	3968	TCAGTTTCGAGTATGGAGCGAAGGTTCCGATCCTGCGCATGTGGCTATCCAGCTGACCT	4027
QY	613	TCTTGCGCTGATGTCCACCGAGGCTCCAGAGACATCACTACCACTGCAAGAACGCG	672
Db	4028	TCTTGCGCTGATGTCCACCGAGGCTCCAGAGACATCACTATCACTGCAAGAACGCG	4087
QY	673	TGCGCTACATGGAACCAAGGCTGGCAACCTCAAGAGGCGCTGCTCCTCAAGGGCTCCA	732
Db	4088	TAGCTTACATGGACCAACAGCTGGCAACCTCAAGAGGCTGCTCCTCAGGGCTCCA	4147
QY	733	ACGAGATCGAGATCCGCGCGGAGGCAACAGCGCTTCACTACAGCGTCACTGTGATG	792
Db	4148	ACGAGATCGAGCTCAGGGCGAAGGCAACAGTCACTTCACTACAGCGCTTGTGATG	4207
QY	793	GCTGCAGAGTCAACCGGAGCTTGGGCAAGACAGTGAATTGAATACAAACCAACCAAGT	852
Db	4208	GCTGCAGAGTCAACCGGAACTTGGGGCAAGACAGTCACTGAATACAAACCAACCAAGA	4267
QY	853	CCTCCCGCTGCCATCATCGATGTGGCCCTTGGACGTTGGTCCCGCAGACAGCAAGT	912
Db	4268	CCTCCCGCTGCCATCATCGATGTGGCTCCTTGGACATGGTGGCCAGACCAAGAT	4327
QY	913	TCGGCTTCGACGTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC	959
Db	4328	TCGGATGGACATTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC	4374

RESULT 15
US-10-784-004-826
; Sequence 826, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 826
; LENGTH: 5721
; TYPE: DNA
; ORGANISM: rat
US-10-784-004-826

Query Match	83.2%;	Score 801.4;	DB 7;	Length 5721;
Best Local Similarity	90.4%;	Pred. No. 3.8e-190;		
Matches 856;	Conservative 0;	Mismatches 91;	Indels 0;	Gaps 0;
QY	13	GATCTAACGGTCTCCCTGGGCGCCCAATTGGGCGCCCTGGTCTCGGGTTCGCACTGGTGATG	72	
Db	3428	GACTCAACGGTCTCCCTGGGCGCCCAATTGGTCTCGGGTTCGCACTGGCGATA	3487	
QY	73	CTGGTCTCTGTGGTCCCGCGGCTCTCGGACCTCTGGTCCCGCTGGTCTCCCGAGCG	132	
Db	3488	GTGGTCTCTGTGGTCCCGCGGACCTCTGGAACCCCTGGGCTCTCCGGTCTCCCGAGCG	3547	
QY	133	CTGGTTTCGACTTCAGGTTCTCTGCCCCAGCCACTCAAGAGAGGGCTCAGGATGGTGCC	192	

Search completed: April 27, 2006, 19:45:09
Job time : 664.548 secs

Db	3548	GTGGTTATGACTTTCAGCTTCTCGCTCAGCCACCTCAGAGAAAGTCTCAAGATGGTGCC	3607
QY	193	GCTACTACCGGGCTGATGATGCAATGTGTGTTCGTGACCGCTGACCTTCGAGGTGCACACCA	252
Db	3608	GTTACTACCGGGCGGATGATGCCAAGTGTGTGTTCGTGACCGCTGACCTTCGAGGTGCACACTA	3667
QY	253	CCCTCAAGAGCTTGAGCCAGCGAGATCGAGAAACATCCGAGAGCCAGAGGAGAGCCGAGAGA	312
Db	3668	CCCTCAAGAGCTTGAGCCAGCGAGATTCAGAAACATCCGAGAGCCCTGAGGGCAGCCGAGAGA	3727
QY	313	ACCCGCGCGCACTTCGCGTGCCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGT	372
Db	3728	ACCCGCGCGCACTTCGCGTGCCTCAAGATGTGCCACTCTGACTGGAAGAGCGGAGAGT	3787
QY	373	ACTGGATTGACCCCAACCAAGGCTGCAACCTTGATGCCATCAAGTCTTCTGCAACATGG	432
Db	3788	ACTGGATCGACCTTACCAAGGCTGCAACCTTGATGCCATCAAGTCTTACTGCAACATGG	3847
QY	433	AGACTGGTGAACCTCGGTGTATCCCACTCAGGCCAGTGTGGCCCAAGAGAACTGGGTACA	492
Db	3848	AGACAGGTTCAGACTGTGTGTTCCTCCACTCAGCCCTCTGTGCTCAGAAAGAACTGGGTACA	3907
QY	493	TCAGCAAGAACCCCAAGGCAACAGAGGATGTCTGGTTTCGGCGAGAGATGACCCGATGGAT	552
Db	3908	TCAGCCCAACCCCAAGGAGAAAGCATGTCTGGTTTGGAGAGAGATGACCCGATGGAT	3967
QY	553	TCAGTTTCGAGTATGGCGGCCAGGGCTCCGACCTGCGCATGTGGCCATCCAGCTGACCT	612
Db	3968	TCAGTTTCGAGTATGGAGCGAAGGTTCCGATCCTGCGCATGTGGCTATCCAGCTGACCT	4027
QY	613	TCTTGCGCTGATGTCCACCGAGGCTTCCAGAAACATCACTTACCACTGCAAGAACGCG	672
Db	4028	TCTTGCGCTGATGTCCACCGAGGCTTCCAGAAACATCACTTATCACTGCAAGAACGCG	4087
QY	673	TGCGCTTACATGGACCAAGGCTGGCAACCTCAAGAGGCGCTGCTCCTCAAGGGCTCCA	732
Db	4088	TAGCTTACATGGACCAACAGACTGGCAACCTCAAGAGGCTGCTCCTCAGGGCTCCA	4147
QY	733	ACGAGATCGAGATCCGCGCGGAGGCAACAGCGCTTCACTACAGCGTCACTGTGATG	792
Db	4148	ACGAGATCGAGCTCAGGGCGAAGGCAACAGTCACTTCACTACAGCAACGCTTGTGATG	4207
QY	793	GCTGCAGAGTCAACCGGAGCTTGGGCAAGACAGTGAATTGAATACAAACCAACCAAGT	852
Db	4208	GCTGCAGAGTCAACCGGAACTTGGGGCAAGACAGTCACTGAATACAAACCAACCAAGA	4267
QY	853	CCTCCCGCTGCCATCATCGATGTGGCCCTTGGACGTTGGTCCCGCAGACAGCAAGT	912
Db	4268	CCTCCCGCTGCCATCATCGATGTGGCTCCTTGGACATGGTGGCCAGACCAAGAT	4327
QY	913	TCGGCTTCGACGTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC	959
Db	4328	TCGGATGGACATTTGGCCCTGTCTGCTTCTGTAAACTCCCTCCATC	4374

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:10:27 ; Search time 81.3728 Seconds
(without alignments)
1679.270 Million cell updates/sec

Title: US-10-677-877a-2
Perfect score: 1707
Sequence: 1 RSNGLPGIPGPRGRGTG.....LDVGAPDQBERGDFGVPCFL 311

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1707	100.0	311	9 AEA28569	Aea28569 Human col
2	1707	100.0	699	9 AEA28581	Aea28581 Human sol
3	1707	100.0	819	9 AEA28573	Aea28573 Human alk
4	1702	99.7	566	9 AEA28577	Aea28577 Human sol
5	1698	99.5	1284	8 ADS98184	AdS98184 Protein f
6	1698	99.5	1464	4 AAB82454	Aab82454 Human pro
7	1698	99.5	1464	5 ABB90764	Abb90764 Human tum
8	1698	99.5	1464	5 ABP68610	Abp68610 Human pan
9	1698	99.5	1464	6 ABUS4471	Abu54471 Human tum
10	1698	99.5	1464	6 ABR47417	AbR47417 Breast ca
11	1698	99.5	1464	6 ABR92064	AbR92064 Human cer
12	1698	99.5	1464	7 ADD14142	Add14142 Human src
13	1698	99.5	1464	7 ADP65246	Adp65246 Human alp
14	1698	99.5	1464	8 ADQ19470	Adq19470 Human sof
15	1698	99.5	1464	8 ADQ29653	Adq29653 Human col
16	1698	99.5	1464	9 ADV87000	Adv87000 Collagen
17	1698	99.5	1464	9 ADV70233	Adv70233 Tumor-ass
18	1698	99.5	1464	9 ADZ26573	Adz26573 Human typ
19	1698	99.5	1464	9 ADZ09874	Adz09874 Human bre
20	1698	99.5	1464	5 AEA04480	Aea04480 Human pol
21	1695	99.3	1461	5 ABG93947	Abg93947 Human pol
22	1695	99.3	1464	2 AAW68485	Aaw68485 Human rec
23	1695	99.3	1464	2 ADD45059	Add45059 Human Pro
24	1695	99.3	1464	7 ADD45055	Add45055 Human Pro

25	1695	99.3	1464	7 ADD45051	Add45051 Human Pro
26	1695	99.3	1464	8 ADR16800	Adr16800 Human col
27	1695	99.3	1464	8 ADR16425	Adr16425 Human col
28	1695	99.3	1464	8 ADR99144	Adr99144 Collagen,
29	1695	99.3	1464	9 ADX58061	Adx58061 Amino aci
30	1691	99.1	366	6 ABR41631	AbR41631 Human DIT
31	1691	99.1	449	3 AAB43439	Aab43439 Human can
32	1691	99.1	465	7 ADE87049	Ade87049 Human pan
33	1691	99.1	886	7 ADE87052	Ade87052 Human pan
34	1691	99.1	1151	7 ADE87050	Ade87050 Human pan
35	1691	99.1	1464	4 AAU14136	Aau14136 Human nov
36	1691	99.1	1464	7 ADE87048	Ade87048 Human pan
37	1691	99.1	1536	7 ADE87051	Ade87051 Human pan
38	1688	98.9	1040	9 ADZ70363	Adz70363 Human pro
39	1676	98.2	1341	2 AAR71701	Aar71701 Collagen
40	1670.5	97.9	1341	3 AAY96122	Aay96122 Collagen
41	1670.5	97.9	1341	5 AAEL16475	Aael16475 Human col
42	1670.5	97.9	1341	5 ABB80733	Abb80733 Collagen
43	1670.5	97.9	1341	5 ABB09625	Abb09625 Amino aci
44	1670.5	97.9	1341	7 ADF13075	Adf13075 Human col
45	1653	96.8	420	5 ABG93946	Abg93946 Deer poly

ALIGNMENTS

RESULT 1
AEA28569
ID AEA28569 standard; protein; 311 AA.

XX AEA28569;

XX 28-JUL-2005 (first entry)

XX Human collagen alpha(I) C-propeptide TO construct protein.

XX protein secretion; fusion protein; protein activation; collagen I.

XX Homo sapiens.

XX Synthetic.

XX Key Location/Qualifiers

XX Region 1..40 /note= "Glycine repeat region"

XX WO2005047850-A2.

XX PD 26-MAY-2005.

XX 04-OCT-2004; 2004WO-US032753.

XX 02-OCT-2003; 2003US-00677877.

XX (GENH-) GENHUNTER CORP.

XX Liang P;

XX WPI; 2005-386406/39.

XX N-PSDB; AEA28568.

XX Generating a secreted trimeric fusion protein by introducing into a cell a DNA construct comprising a promoter linked to a template encoding a signal peptide sequence followed by in-frame fusion to polypeptide to be trimerized.

XX Claim 11; SEQ ID NO 2; 47pp; English.

XX The invention relates to a novel method for generating a secreted trimeric fusion protein. The method comprises creating a DNA construct comprising a transcriptional promoter linked to a template encoding a signal peptide sequence followed, via an in-frame fusion, to the polypeptide to be trimerized, which in turn is joined in-frame to a polypeptide capable of self-trimerization which is heterologous to the

CC higher affinity to their ligand than their dimeric counterparts. Such
CC rationally designed soluble trimeric receptor analogs could significantly
CC increase clinical benefit, as well as lower the amount or frequency of
CC drug injections for each patient. The method of the invention may be
CC useful for generating a secreted trimeric fusion protein to be used for
CC more efficient neutralization of the biological activities of their
CC naturally occurring trimeric ligands. The current sequence is that of the
CC human soluble TNF-R1I-human collagen alpha(I) C-propeptide T0 fusion
CC protein (sTNF-R1I-T0) of the invention. The collagen T0 construct
CC includes a partial glycine-repeat triple helical region.

XX
SQ Sequence 566 AA;

Query Match 99.7%; Score 1702; DB 9; Length 566;
Best Local Similarity 100.0%; Pred. No. 3.8e-134; Indels 0; Gaps 0;
Matches 310; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SNGLPGPIGPPGRGTGDAGVPGPPGPPGPPPSAGDFSLPQPQKAHDGGR 61
DB |||||
QY 257 SNGLPGPIGPPGRGTGDAGVPGPPGPPGPPPSAGDFSLPQPQKAHDGGR 316
DB |||||
QY 62 YRADDANVVRDRDLEVDVTLKSLSQIENIRSPGSRKNPARTCDLKNCHSDWKSGEY 121
DB |||||
QY 317 YRADDANVVRDRDLEVDVTLKSLSQIENIRSPGSRKNPARTCDLKNCHSDWKSGEY 376
DB |||||
QY 122 WIDPNOGNCNLDAIKVFCNMTGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGF 181
DB |||||
QY 377 WIDPNOGNCNLDAIKVFCNMTGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGF 436
DB |||||
QY 182 QFYGGGSDPADVAIQLTFLRLMSTASQNIYTHCKNSVAYMDQQTGNLKKALLKGSN 241
DB |||||
QY 437 QFYGGGSDPADVAIQLTFLRLMSTASQNIYTHCKNSVAYMDQQTGNLKKALLKGSN 496
DB |||||
QY 242 EIRAEGRNRFYTSVTDGCTSHTGAWGKTVEYTKTSRRLPIIDVAPLDVGAPDQEF 301
DB |||||
QY 497 EIRAEGRNRFYTSVTDGCTSHTGAWGKTVEYTKTSRRLPIIDVAPLDVGAPDQEF 556
DB |||||

RESULT 5

AD98184

ID AD98184 standard; protein; 1284 AA.

XX AC

AD98184;

XX 30-DEC-2004 (first entry)

XX Protein factor discovery related isolated human polypeptide, SEQ ID 448.

XX antiinflammatory; cytostatic; antimicrobial; gene therapy; inflammation;

XX leukaemia; nervous system disorder; infection.

XX Homo sapiens.

XX WO2004087874-A2.

XX 14-OCT-2004.

XX 24-MAR-2004; 2004WO-US009202.

XX 28-MAR-2003; 2003US-0458824P.

XX (NUVE-) NUVELO INC.

XX (DRMA/) DRMANAC R T.

XX Tang YT, Zhou P, Wang J, Wang ZW, Hu T;

XX WPI; 2004-737686/72.

XX N-PSDB; ADS97949.

XX

PT New polynucleotides encoding a polypeptide with biological activity,
PT useful for treating inflammation, leukemias, nervous system disorders, or
PT infections.

PS Claim 20; SEQ ID NO 448; 253pp; English.

XX The invention relates to a novel isolated polynucleotide comprising any
CC of the 235 nucleotide sequences described in the specification. The
CC invention further comprises: an isolated polynucleotide encoding a
CC polypeptide with biological activity, where the polynucleotide hybridizes
CC to one of the 235 novel polynucleotides under stringent hybridization
CC conditions, or having greater than about 99% sequence identity with the
CC novel polynucleotide; a vector comprising a novel polynucleotide; an
CC expression vector comprising the novel polynucleotide; a host cell
CC genetically engineered to comprise the novel polynucleotide, which can be
CC operatively associated with a regulatory sequence that modulates
CC expression of the polynucleotide in the host cell; an isolated
CC polypeptide encoded by the novel polynucleotide, or a polynucleotide
CC hybridizing under stringent conditions to the novel polynucleotide; a
CC composition comprising the polypeptide and a carrier; an antibody
CC directed against the polypeptide; a method for detecting the novel
CC polynucleotide in a sample; a method for detecting the polypeptide in a
CC sample; a method for identifying a compound that binds to the polypeptide
CC; a method for producing the polypeptide; an isolated polypeptide
CC comprising any of the 235 amino acid sequences described in the
CC specification; and a collection of polynucleotides comprising of at least
CC one of the polynucleotides cited above. The polypeptides and
CC polynucleotides of the invention have antiinflammatory, cytostatic, and
CC antimicrobial activities. The novel polynucleotide may be used to treat
CC disorders by gene therapy. The polypeptides and polynucleotides are
CC useful for treating inflammation, leukemias, nervous system disorders,
CC or infections. This sequence represents one of the 235 novel isolated
CC polypeptides of the invention.

XX SQ Sequence 1284 AA;

Query Match 99.5%; Score 1698; DB 8; Length 1284;

Best Local Similarity 100.0%; Pred. No. 2.5e-133;

Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NGLPGPIGPPGRGTGDAGVPGPPGPPGPPPSAGDFSLPQPQKAHDGGRY 62

DB |||||

QY 976 NGLPGPIGPPGRGTGDAGVPGPPGPPGPPPSAGDFSLPQPQKAHDGGRY 1035

DB |||||

QY 63 YRADDANVVRDRDLEVDVTLKSLSQIENIRSPGSRKNPARTCDLKNCHSDWKSGEY 122

DB |||||

QY 1036 YRADDANVVRDRDLEVDVTLKSLSQIENIRSPGSRKNPARTCDLKNCHSDWKSGEY 1095

DB |||||

QY 123 IDPNOGNCNLDAIKVFCNMTGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQ 182

DB |||||

QY 1096 IDPNOGNCNLDAIKVFCNMTGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQ 1155

DB |||||

QY 183 FRYGGGSDPADVAIQLTFLRLMSTASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNE 242

DB |||||

QY 1156 FRYGGGSDPADVAIQLTFLRLMSTASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNE 1215

DB |||||

QY 243 IEIRAEGRNRFYTSVTDGCTSHTGAWGKTVEYTKTSRRLPIIDVAPLDVGAPDQEF 302

DB |||||

QY 1216 IEIRAEGRNRFYTSVTDGCTSHTGAWGKTVEYTKTSRRLPIIDVAPLDVGAPDQEF 1275

DB |||||

QY 303 FDVGVPVCF 311

DB 1276 FDVGVPVCF 1284

RESULT 6

AAB82454

ID AAB82454 standard; protein; 1464 AA.

XX AC

AAB82454;

XX 22-AUG-2001 (first entry)

XX

Db 1456 FDVGPVCFL 1464

|||||

RESULT 7

ABB90764

ID ABB90764 standard; protein; 1464 AA.

XX AC ABB90764;

XX AC

DT 30-MAY-2002 (first entry)

XX AC

XX DE Human Tumour Endothelial Marker polypeptide SEQ ID NO 261.

XX DE

XX KW Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostatic;

XX KW normal endothelial marker; pan-endothelial marker; immunostimulant;

XX KW antiangiogenic; tumour; neoangiogenesis; vascularised tumour;

XX KW polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis;

XX KW psoriasis.

XX OS Homo sapiens.

XX OS

XX WN W0200210217-A2.

XX EN

XX PD 07-FEB-2002.

XX PD

XX PF 01-AUG-2001; 2001WO-US024031.

XX PF

XX PR 02-AUG-2000; 2000US-0222599P.

XX PR

XX PR 11-AUG-2000; 2000US-0224360P.

XX PR

XX PR 11-APR-2001; 2001US-0282850P.

XX PR

XX PA (UYJO) UNIV JOHNS HOPKINS.

XX PA

XX PI St Croix B, Kinzler KW, Vogelstein B;

XX PI

XX DR WPI; 2002-291856/33.

XX DR

XX DR N-PSDB; ABL92119.

XX DR

XX PT An isolated molecule comprising an antibody variable region which

XX PT specifically binds to an extracellular domain of a tumor endothelial

XX PT marker (TEM) protein, useful for inhibiting tumor growth.

XX PT

XX PS Claim 54; Page 265-268; 331pp; English.

XX PS

XX CC The invention relates to an isolated molecule comprising an antibody

XX CC variable region which specifically binds to an extracellular domain of a

XX CC tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,

XX CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM

XX CC proteins have cytostatic, immunostimulant and antiangiogenic activity.

XX CC They are useful for inhibiting tumour growth, neoangiogenesis in subjects

XX CC bearing a vascularised tumour, polycystic kidney disease, diabetic

XX CC retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM

XX CC genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)

XX CC are disclosed, as are marker oligonucleotide sequences: tumour

XX CC endothelial markers (TEM) ABL91996-ABL92041 and ABL92143-ABL92191; normal

XX CC endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers

XX CC (PEM) ABL91903-ABL91995

XX CC

SQ Sequence 1464 AA;

Query Match 99.5%; Score 1698; DB 5; Length 1464;

Best Local Similarity 100.0%; Pred. No. 3e-133;

Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NGLPGPIGPPGRGTGDAGFVGP GPPGPPGPPPSAGFDFSLFPQPEKAHDGGRY 62

|||||

Db 1156 NGLPGPIGPPGRGTGDAGFVGP GPPGPPGPPPSAGFDFSLFPQPEKAHDGGRY 1215

|||||

QY 63 YRADDANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMSCHDWKSGEYW 122

|||||

Db 1216 YRADDANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMSCHDWKSGEYW 1275

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:14:30 ; Search time 28.4247 Seconds
(without alignments)
1052.725 Million cell updates/sec

Title: US-10-677-877A-2
Perfect score: 1707
Sequence: 1 RSNGLPGIPGPPGRGRTGD.....LDVGAPDQERFDVGPVCFLL 311

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : 1: PIR.80:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1688	98.9	1464	1 CGHUI1	collagen alpha 1(I
2	1591	93.2	473	2 I50629	collagen - chicken
3	1564	91.6	1453	2 S21626	collagen alpha 1(I
4	1213.5	71.1	1487	1 CGHU6C	collagen alpha 1(I
5	1212.5	71.0	1418	2 T45467	collagen alpha 1(I
6	1206.5	70.7	1419	2 A41182	collagen alpha 1(I
7	1206.5	70.7	1487	2 B41182	collagen alpha 1(I
8	1177.5	69.0	464	2 S59513	collagen II A1 pro
9	1175	68.8	1486	1 B40333	collagen alpha 1(I
10	1159.5	67.9	1492	2 A40333	collagen alpha 1(I
11	1124.5	65.9	288	1 CGCH8C	collagen alpha 1(I
12	1067	62.5	964	1 CGCH2S	collagen alpha 2(I
13	1066	62.4	365	2 S10847	collagen alpha 2(I
14	1019	59.7	1366	1 CGHU2S	collagen alpha 2(I
15	1007	59.0	1466	1 CGHU7L	collagen alpha 1(I
16	1006.5	59.0	615	2 A05269	collagen alpha 1(I
17	1003	58.8	636	2 S41067	collagen alpha 1(I
18	1001.5	58.7	1464	2 S59856	collagen alpha 1(I
19	980	57.4	323	2 A61396	collagen alpha 1(I
20	974.5	57.1	1373	1 A43291	collagen alpha 2(I
21	909.5	53.3	1496	1 CGHU2V	collagen alpha 2(V
22	899.5	52.7	1497	2 I49607	procollagen type V
23	786.5	46.1	673	1 CGBO8C	collagen alpha 1(I
24	663	38.8	143	2 S07234	collagen alpha 1(I
25	643	37.7	632	2 S42731	collagen alpha 1 c
26	642	37.6	730	2 A36226	collagen alpha 1 c
27	642	37.6	1414	1 S23809	collagen alpha 2(I
28	639	37.4	1806	1 CGHUI1	collagen alpha 1(X
29	638	37.4	1843	2 S18803	collagen alpha 1(V

30	636	37.3	1838	1 CGHUIV	collagen alpha 1(V
31	627	36.7	888	2 S28791	collagen alpha 1(X
32	611	35.8	3198	2 A43426	collagen alpha 2 f
33	608	35.6	482	2 B31795	collagen alpha 1(X
34	537.5	31.5	1051	2 A35763	collagen alpha 2 c
35	521	30.5	193	2 S07133	collagen alpha 1(I
36	462	27.1	547	2 A36046	collagen alpha cha
37	430.5	25.2	1546	1 CGHU2E	collagen alpha 2(X
38	353	20.7	91	2 S71545	alpha 2(I) procoll
39	334.5	19.6	779	1 CGBOIS	collagen alpha 1(I
40	306.5	18.0	671	1 CGRTIS	collagen alpha 1(I
41	262	15.3	69	2 I46677	collagen alpha 2(I
42	262	15.3	1042	1 CGCHIS	collagen alpha 1(I
43	203	11.9	39	2 I65269	collagen alpha 1(I
44	176.5	10.3	102	2 B34770	ORF2 protein - sai
45	173	10.1	1747	2 A45974	collagen alpha 1(X

ALIGNMENTS

RESULT 1

CGHUI1

collagen alpha 1(I) chain precursor - human

N;Alternate names: procollagen alpha 1(I) chain

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1981 #sequence_revision 04-Oct-1996 #text change 31-Dec-2004

C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11

5269; A29439; I53466; A02852; I37247

R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.

Gene 67, 105-115, 1988

A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five e

A;Reference number: I60114; MUID:88329734; PMID:2843432

A;Accession: I60114

A;Status: translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-369, 'L', 371-589 <DAL>

A;Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP

R;Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock

Biochem. J. 253, 919-922, 1988

A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human

A;Reference number: S01143; MUID:89025644; PMID:3178743

A;Accession: S01143

A;Molecule type: mRNA

A;Residues: 1-472 <TRO>

A;Cross-references: UNIPARC:UPI000016A6F9; EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID

A;Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;

Nature 310, 337-340, 1984

A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of

A;Reference number: A93335; MUID:84270697; PMID:6462220

A;Accession: A93335

A;Molecule type: DNA

A;Residues: 1-58, 'Q', 60-181 <CHU>

A;Cross-references: UNIPARC:UPI000017B3C; EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID

R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.O.; Bernard, M.P.; Ramirez, F.; de Wet, W.

J. Biol. Chem. 262, 15151-15157, 1987

A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene en

A;Reference number: I55254; MUID:88033098; PMID:2822714

A;Accession: I55254

A;Status: translation not shown; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-45 <ROS>

A;Cross-references: UNIPARC:UPI000016A6B2; GB:J02829; NID:g180387; PIDN:AAAS1993.1; PID:5

R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 8869-8873, 1987

A;Title: Regulatory elements in the first intron contribute to transcriptional control of

A;Reference number: A39943; MUID:88097389; PMID:3480516

A;Accession: A39943

A;Molecule type: DNA

A;Residues: 1-34 <BOR>

A;Cross-references: UNIPARC:UPI000016A6B1; GB:J03559; NID:g180876; PIDN:AAAS2052.1; PID:5

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter e
A;Reference number: I55237; MUID:85130970; PMID:2857713
A;Accession: I55237
A;Status: translation not shown; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-34 <CH2>
A;Cross-references: UNIPARC:UPI000016A6B1; GB:M10627; NID:G180383; PIDN:AAA51992.1; PID:
R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollist
J. Biol. Chem. 265, 6312-6317, 1990
A;Title: In vivo and in vitro noncovalent association of excised alpha(I) amino-termina
rome, type VII.
A;Reference number: A35233; MUID:90202908; PMID:2318855
A;Accession: A35233
A;Molecule type: protein
A;Residues: 33-52 <WR>
A;Cross-references: UNIPARC:UPI0000173B3D
A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved
R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F.
EMBO J. 8, 1705-1710, 1989
A;Title: A base substitution in the exon of a collagen gene causes alternative splicing
A;Reference number: S09400; MUID:89356643; PMID:2767050
A;Accession: S09400
A;Molecule type: mRNA
A;Residues: 156-183 <WEI>
A;Cross-references: UNIPARC:UPI0000173B3E
R;Click, E.M.; Bornstein, P.
Biochemistry 9, 4699-4706, 1970
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha
A;Reference number: A90567; MUID:71038625; PMID:5529814
A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A;Accession: B90567
A;Molecule type: protein
A;Residues: 182-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z'
A;Cross-references: UNIPARC:UPI0000173B3F
A;Experimental source: skin
A;Note: evidence for 170-allysine
R;Baetge, B.; Norbom, H.; Diebold, J.; Lehmann, H.; Bodo, M.; Deutzmann, R.; Mueller, H
Eur. J. Biochem. 192, 153-159, 1990
A;Title: A critical crosslink region in human-bone-derived collagen type I. Specific cle
A;Reference number: S11372; MUID:90382436; PMID:2169412
A;Accession: S11372
A;Molecule type: protein
A;Residues: 175-187, 274-287, 'P', 289 <BAE>
A;Cross-references: UNIPARC:UPI0000173B40; UNIPARC:UPI0000173B41
R;Deak, S.B.; Scholz, P.M.; Amenta, P.S.; Constantinou, C.D.; Levi-Minzi, S.A.; Gonzalez
J. Biol. Chem. 266, 21827-21832, 1991
A;Title: The substitution of arginine for glycine 85 of the alpha 1(I) procollagen chain
operative melting of intact type I collagen.
A;Reference number: I55342; MUID:92042092; PMID:1718984
A;Accession: I55342
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 258-268, 1347-1357 <DEA>
A;Cross-references: UNIPARC:UPI0000173B42; UNIPARC:UPI0000173B43; GB:S67495; NID:G239007
A;Note: sequences from the 5' and 3' ends only are shown; mutant sequence 263-Arg report
R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.
J. Biol. Chem. 245, 5042-5048, 1970
A;Title: Comparative study of glycopeptides derived from selected vertebrate collagens.
A;Reference number: A92069; MUID:71001508; PMID:4319110
A;Accession: A92069
A;Molecule type: protein
A;Residues: 263-268 <MOR>
A;Cross-references: UNIPARC:UPI000014DF11
A;Experimental source: skin
A;Note: attachment of 2-O-alpha-D-glucosyl-O-beta-D-galactose to 5-hydroxylysine
R;Labhard, M.E.; Hollister, D.W.
Matrix 10, 124-130, 1990
A;Title: Segmental amplification of the entire helical and telopeptide regions of the ch
A;Reference number: S15989; MUID:90326017; PMID:2374517
A;Accession: S15989
A;Molecule type: mRNA

A;Residues: 281-302, 402-420, 823-843, 925-944, 1026-1045, 1143-1162 <LAB>
A;Cross-references: UNIPARC:UPI0000173B44; UNIPARC:UPI0000173B45; UNIPARC:UPI0000173B46;
R;Wirtz, M.K.; Rao, V.H.; Glanville, R.W.; Labhard, M.E.; Pretorius, P.J.; de Vries, W.N
Connect. Tissue Res. 29, 1-11, 1993
A;Title: A cysteine for glycine substitution at position 175 in an alpha 1 (I) chain of t
A;Reference number: I52905; MUID:93339042; PMID:8339541
A;Accession: I52905
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 342-352, 'C', 354-359 <WI2>
A;Cross-references: UNIPARC:UPI0000070EB6; GB:S64717; NID:G408195; PIDN:AA827677.1; PID:
R;Bernard, M.P.; Chu, M.L.; Myers, J.C.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.
Biochemistry 22, 5213-5223, 1983
A;Title: Nucleotide sequences of complementary deoxyribonucleic acids for the proalpha 1 (I)
A;Reference number: A90476; MUID:84080385; PMID:6689127
A;Accession: A90476
A;Molecule type: mRNA
A;Residues: 425-1250, 'X', 1252-1328, 'S', 1330-1390, 'X', 1392-1464 <BER>
A;Cross-references: UNIPARC:UPI0000173B4A; GB:K01228; NID:G180391; PIDN:AAA51995.1; PID:
R;Chu, M.L.; Gargiulo, V.; Williams, C.J.; Ramirez, F.
J. Biol. Chem. 260, 691-694, 1985
A;Title: Multixon deletion in an osteogenesis imperfecta variant with increased type III
A;Reference number: A22161; MUID:85104934; PMID:2981843
A;Accession: A22161
A;Molecule type: DNA
A;Residues: 472-594, 'R', 596-607 <CH3>
A;Cross-references: UNIPARC:UPI000011F796; GB:K03178; GB:K03179; NID:G179612; NID:G17961;
A;Note: the authors translated the codon CGT for residue 595 as Pro
R;Wallis, G.A.; Starman, B.J.; Zinn, A.B.; Byers, P.H.
Am. J. Hum. Genet. 46, 1034-1040, 1990
A;Title: Variable expression of osteogenesis imperfecta in a nuclear family is explained
A;Reference number: A35336; MUID:90352792; PMID:2339700
A;Accession: A35336
A;Molecule type: mRNA
A;Residues: 710-720, 'E', 722-737, 'E', 739-745 <WAL>
A;Cross-references: UNIPARC:UPI0000173B4B
A;Note: the authors translated the codons CAG for 721 and CGT for 738 as Glu
R;Forlino, A.; Zolerzi, F.; Valli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes,
Hum. Mol. Genet. 3, 2201-2206, 1994
A;Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the cel
A;Reference number: I54365; MUID:95187161; PMID:7881420
A;Accession: I54365
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 746-766, 'S', 768-781 <FOR>
A;Cross-references: UNIPARC:UPI000016A6FA; GB:I47667; NID:G1009093; PIDN:AA859576.1; PID
R;Chessler, S.D.; Wallis, G.A.; Byers, P.H.
J. Biol. Chem. 268, 18218-18225, 1993
A;Title: Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I) chain of ty
A;Reference number: A47426; MUID:93352646; PMID:8349697
A;Accession: A47426
A;Molecule type: mRNA
A;Residues: 1179-1276, 'H', 1278-1336, 1339-1387, 'R', 1389-1464 <CHE>
A;Cross-references: UNIPARC:UPI0000073A2A; GB:S64596; NID:G407599; PIDN:AA827856.1; PID:
A;Note: sequence extracted from NCBI backbone (NCBI:136444, NCBI:136445)
A;Note: does not represent an experimentally determined sequence but three different mut
A;Accession: B47426
A;Molecule type: mRNA
A;Residues: 1179-1464 <CH4>
A;Cross-references: UNIPARC:UPI0000173B4C
A;Experimental source: normal dermal fibroblast culture
A;Accession: C47426
A;Molecule type: mRNA
A;Residues: 1179-1276, 'H', 1278-1464 <CH5>
A;Cross-references: UNIPARC:UPI0000173B4D
A;Experimental source: fetal cell 86-237
A;Accession: D47426
A;Molecule type: mRNA
A;Residues: 1179-1336, 1339-1464 <CH6>
A;Cross-references: UNIPARC:UPI0000173B4E

Query Match	98.9%	Score 1688;	DB 1;	Length 1464;
Best Local Similarity	99.0%	Pred. No. 8.4e-117;		
Matches 306; Conservative	3;	Mismatches 0;	Indels 0;	Gaps 0;

RESULT 3

C;Species: Mus musculus (house mouse)
C;Date: 13-Jan-1995 #sequence revision 25-Apr-1997 #text change 09-Jul-2004
C;Accession: S57243; S16374; A23982; I49559; I49557; S39789; I48300; S21626
E;Li, S.-W.; Khillan, J.; Prockop, D.J.

R;Li, S.W.; Khillan, J.; Prockop, D.J.
Matrix Biol. 14, 593-595, 1994
A;Title: The complete cDNA coding sequence
A;Reference number: S57243

A;Reference number: S57243
A;Accession: S57243
A;Molecule type: mRNA
A;Residues: 1-1453 <LIS>

A; Residues: 1-1453 <LIS>
A; Cross-references: UNIPROT:P11087; UNIPARC:UPI0000027558; E
R; Metaeasranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991

Biochim. Biophys. Acta 1089, 241-243, 1991
 A; Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.
 A; Reference number: S16176; MUID: 91274355; PMID: 2054384
 A; Accession: S16374
 A; Status: preliminary

A;ACCESSION: S16374
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1442-1453 <MET>
A;Cross-references: UNIPARC

A: Accession: U01421 (EMBL)
A: Cross-references: UNIPARC:UPI0000000691; EMBL:X57981.1
R: French, B.T.; Lee, W.H.; Maul, G.G.
Gene 39, 311-312, 1985
A: Title: Nucleotide sequence of a cDNA clone for mouse

A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.
A;Reference number: A23982; MUID:86137403; PMID:3841523
A;Accession: A23982
A;Molecule type: mRNA

A;Molecule type: mRNA
A;Residues: 518-1128 <FRE>
A;Cross-references: UNIPARC:UPI000016CC46; G
R;Monson, J.M.; Friedman, J.; McCarthy, B.J.

R; Monson, J.M.; Friedman, J.; McCarthy, B.J.
Mol. Cell. Biol. 2, 1362-1371, 1982
A; Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for
A; Reference number: 149559; MUID: 83141374; PMID: 6298597

A:Accession: I49559
A:Reference number: I49559; MUID:83141374; PMID:6298559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A; Molecule type: DNA
A; Residues: 735-1130 <RES>
A; Cross-references: UNIPARC
R; Harbers, K.; Kuehn, M.; D
Proc Natl Acad Sci U S

Query Match q3 2% Score 1591: DB 2: Length 473:
F;244-473/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
A;cross-references: UNIPROT:P2425; UNIPARC:U0400001125; EMBL:U00401;
P;244-473/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 93.2%; Score 1591; DB 2; Length 473;
Best Local Similarity 90.3%; Pred. No. 3.1e-110;
Matches 279; Conservative 21; Mismatches 9; Indels 0; Gaps 0;

QY 3 NGLPGPIGPGRGTGDAGVPGPPGPPGPPSAGDFSLPPOPEKAHDGGRY 62
|||||
Dd 165 NGLPGPIGPGRGTGEVVGPPGPPGPPSGGFSLPPOPEKAHDGGRY 222
|||||

[illegible]

```

db      225  YRADDANVRDRDLEVDTTLSLSOOIENIRSPETRKNPARTCDLCKCHGDMKSGEYW 28
QY      123  IDPNQCNLDIAKVFNCMEGTGYPTQPSVAQKNWTYISKPKOKRHHVNGESMTDGFQ 18

```

Qy	Db	Qy	Db
123	185	123	185
IDPQNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYLSKNPKDKRHVWFGESMTDGFQ	IDPQNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYLSKNPKDKRHVWFGESMTDGFQ	IDPQNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYLSKNPKDKRHVWFGESMTDGFQ	IDPQNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYLSKNPKDKRHVWFGESMTDGFQ
285	344	285	344
IDPQNCNLDAIKVFCNMETGETCVYPTQATTAQKNWYLSKNPKDKGHVWFGESMTDGFQ	IDPQNCNLDAIKVFCNMETGETCVYPTQATTAQKNWYLSKNPKDKGHVWFGESMTDGFQ	IDPQNCNLDAIKVFCNMETGETCVYPTQATTAQKNWYLSKNPKDKGHVWFGESMTDGFQ	IDPQNCNLDAIKVFCNMETGETCVYPTQATTAQKNWYLSKNPKDKGHVWFGESMTDGFQ

QY 183 FEYGGGSDPADVAIQLTFLRLMSTEASQNTYTHCKNSVAYMDQQTGNLKKALLKGSNE 24

A/Accession: I48300
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-80, 'E', '82-105, 'D', 107-147 <RSF>
A/Cross-references: UNIPARC:UPI000017738B; EMBL:X54876; NID:g50486; PIDN:CAA38657.1; PID:9180884
C/Genetics:
A/Gene: COL1A1
A/Intons: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3
C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
C/Keywords: coiled coil; extracellular matrix; glycoprotein; heterotrimer; triple helix
F/1-22/Domain: signal sequence #status predicted <SIG>
F/23-151/Domain: amino-terminal propeptide #status predicted <PRO>
F/30-89/Domain: von Willebrand factor type C repeat homology <WRC>
F/152-1453/Product: collagen alpha 1(I) chain #status predicted <MAT>
F/1224-1453/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match	91.6%	Score 1564	DB 2	Length 1453
Best Local Similarity	89.6%	Pred. No. 1.3e-107		
Matches 277	Conservative 15	Mismatches 17	Indels 0	Gaps 0
QY 3	NGLPGP	IGPPRGRTGDAGVGP	PGPPGPPGPPSAGFDSFLPQPQEKAHDGGRY 62	
Db				
QY 63	YRADDANVVRDRLEVD	TTLKSLSQIENIRSPGSRKNPARTCRDLKWC	HSDMKSGEY 122	
Db				
QY 123	IDNQGNLDAIKVFC	MMETGETCVPTQPSVAQKNWYISKPKDKRHWFGESMTDGFQ 182		
Db				
QY 1265	IDNQGNLDAIKVYCN	MMETGQTCTVPTQPSVPQKNWYISPNPKKKHVMFGESMTDGF 1324		
Db				
QY 183	FEYGGGSDPADVAI	OLTFLRLMSTRASQNTVHCNNSVAYMDQQTGNLKKALLKSGNS 242		
Db				
QY 1325	FEYGGGSDPTDAI	OLTFLRLMSTRASQNTVHCNNSVAYMDQQTGNLKKALLKSGNS 1384		
Db				
QY 243	IEIRAGNSRFTYSVT	VDGCTSHGTGAWGKTIVYKTKTSRLPIIDVAPLIDVAGAPQDFG 302		
Db				
QY 303	FDVGPVCF	L 311		
Db				
Db 1445	LDIGPACFV	1453		
RESULT 4				
CGHU6C				
N/Alternate names:	collagen alpha 1(II) chain precursor [validated] - human			
C/Species:	Homo sapiens (man)			
C/Date:	28-May-1986 #sequence revision 01-Sep-1995 #text change 31-Dec-2004			
C/Accession:	A38513; S06715; S24270; A24828; S06496; A35428; A30147; A33116; S64674; S637250; I37251; I37252; I37253; I37254; I55338; I59535; I61910			
R/Ryan, M.C.;	Sieraski, M.;	Sandell, L.J.		
Genomics 8,	41-48,	1990		
A/Title:	The human type II procollagen gene: identification of an additional protein-cod			
A/Reference number:	A38513; MUID:91184811; PMID:2081599			
A/Accession:	A38513			
A/Molecule type:	DNA			
A/Residues:	1-103 <RVA>			
A/Cross-references:	UNIPROT:P02458; UNIPROT:Q14042; UNIPROT:Q16672; UNIPROT:Q12985; UNIP			
IG180884				
R/Su, M.W.;	Lee, B.;	Ramirez, F.;	Machado, M.;	Horton, W.
Nucleic Acids Res.	17,	9473,	1989	
A/Title:	Nucleotide sequence of the full length cDNA encoding for human type II procolla			
A/Reference number:	S06715; MUID:90067946; PMID:2587267			
A/Accession:	S06715			
A/Molecule type:	mRNA			
A/Residues:	1-28, 'R', '99-1487 <SU2>			
A/Cross-references:	UNIPARC:UPI0000126D15; EMBL:X16468; NID:g29515; PIDN:CAA34488.1; PID			
A/Note:	alternative splice form 1			
R/Vikkula, M.;	Metsaeranta, M.;	Syvaenen, A.C.;	Ala-Kokko, L.;	Vuorio, E.;
Pelttonen, L.				

Biochem. J. 285, 287-294, 1992
A/Title: Structural analysis of the regulatory elements of the type-II procollagen gene.
A/Reference number: S24270; MUID:92344585; PMID:1637314
A/Accession: S24270
A/Status: translation not shown
A/Molecule type: DNA
A/Residues: 1-28 <VIK>
A/Cross-references: UNIPARC:UPI0000173B64; EMBL:X58709; GB:S40537; NID:g35659
A/Note: this translation is not annotated in GenBank entry HSPROCOB1, release 111.0
R/Nunez, A.M.; Kohno, K.; | Martin, G.R.; | Yamada, Y. | || Gene 44, | 11-16, | 1986 | | |
A/Title:	Promoter region of the human pro-alpha-1-(II)-collagen gene.			
A/Reference number:	A24828; MUID:87031574; PMID:30211582			
A/Accession:	A24828			
A/Molecule type:	DNA			
A/Residues:	1-8, 'T', '10-28 <NUN>			
A/Cross-references:	UNIPARC:UPI000016A71A; GB:M25698; NID:G180872; PIDN:AAAS2051.1; PID:9			
R/Baldwin, C.T.;	Reginato, A.M.;	Smith, C.;	Jimenez, S.A.;	Prockop, D.J.
Biochem. J.	262,	521-528,	1989	
A/Title:	Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)			
A/Reference number:	S06496; MUID:90026318; PMID:2803268			
A/Accession:	S06496			
A/Molecule type:	mRNA			
A/Residues:	7-26, 'R', '99-157, 'P', '159-440, 'G', '442-456, 'E', '458-640, 'A', '642-831, 'PA', '834, 'P',			
A/Cross-references:	UNIPARC:UPI0000173B65; EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:			
A/Note:	alternative splice form 1			
R/Ryan, M.C.;	Sandell, L.J.			
J. Biol. Chem.	265,	10334-10339,	1990	
A/Title:	Differential expression of a cysteine-rich domain in the amino-terminal propepti			
A/Reference number:	A35428; MUID:90285153; PMID:2355003			
A/Accession:	A35428			
A/Status:	not compared with conceptual translation			
A/Molecule type:	mRNA			
A/Residues:	27-81, 'L', '83-103 <RYA2>			
A/Cross-references:	UNIPARC:UPI0000173B66			
A/Note:	alternative splice form 2; splicing appears to be under developmental regulation			
R/Su, M.W.;	Benson-Chanda, V.;	Vissing, H.;	Ramirez, F.	
Genomics 4,	438-441,	1989		
A/Title:	Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide confi			
A/Reference number:	A30147; MUID:89233138; PMID:2714801			
A/Accession:	A30147			
A/Molecule type:	DNA			
A/Residues:	104-157, 'P', '159-236 <SUM>			
A/Cross-references:	UNIPARC:UPI0000173B67; GB:J03065; GB:M23660; GB:M25655; GB:M25656; GB:			
R/Ala-Kokko, L.;	Baldwin, C.T.;	Moskowitz, R.W.;	Prockop, D.J.	
proc. Natl. Acad. Sci. U.S.A.	87,	6565-6568,	1990	
A/Title:	Single base mutation in the type II procollagen gene (COL2A1) as a cause of prin			
A/Reference number:	A94227; MUID:90370826; PMID:1975693			
A/Accession:	A33116			
A/Molecule type:	DNA			
A/Residues:	171-172, 'C', '174-175 <ALA>			
A/Cross-references:	UNIPARC:UPI0000173B68			
A/Note:	mutant sequence from a family with family with primary generalized osteoarthritis			
R/Diab, M.;	Wu, J.J.;	Eyre, D.R.		
Biochem. J.	314,	327-332,	1996	
A/Title:	Collagen type IX from human cartilage: a structural profile of intermolecular cr			
A/Reference number:	S64673; MUID:96195147; PMID:8660302			
A/Accession:	S64674			
A/Molecule type:	protein			
A/Residues:	188-189, 'X', '191-195;1224-1230, 'X', '1232-1236 <DIA>			
A/Cross-references:	UNIPARC:UPI0000173B69; UNIPARC:UPI0000173B6A			
R/Franc, S.;	Marzin, E.;	Boutillon, M.M.;	Lafont, R.;	Lechene de la Porte, P.;
Eur. J. Biochem.	234,	125-131,	1995	
A/Title:	Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil cart			
A/Reference number:	S63514; MUID:96096730; PMID:8529631			
A/Accession:	S63514			
A/Molecule type:	protein			
A/Residues:	243-261;575-590;756-763, 'X', '765-779 <FRA>			
A/Cross-references:	UNIPARC:UPI0000173B6B; UNIPARC:UPI0000173B6C; UNIPARC:UPI0000173B6D			
R/Tiller, G.E.;	Weis, M.A.;	Polumbo, P.A.;	Gruber, H.E.;	Rimoin, D.L.;
Am. J. Hum. Genet.	56,	388-395,	1995	
A/Title:	An RNA-splicing mutation (G>SIVS20) in the type II collagen gene (COL2A1) in a			
A/Reference number:	I38867; MUID:95150028; PMID:7847372			

C/Species: Brachydanio rerio (zebra fish)
C/Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 09-Jul-2004
C/Accession: S59513
R/Yan, Y.; Hotta, K.; Riggleman, B.; Postlethwait, J.H.
submitted to the EMBL Data Library, March 1995
A/Description: Expression of a type II collagen gene in the zebrafish embryonic axis.
A/Reference number: S59513
A/Accession: S59513
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-464 <YAN>
A/Cross-references: UNIPROT:Q90412; UNIPARC:UPI00000FC847; EMBL:U23822; NID:q773660;
C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F/236-464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 69.0%; Score 1177.5; DB 2; Length 464;
Best Local Similarity 68.2%; Pred. No. 1.2e-79;
Matches 214; Conservative 37; Mismatches 56; Indels 7; Gaps 4;

Qy 2 SNGLPFGIPGPPGRGTGAGVPGPPGPPGPPSAGFD---FSFLPQPPEKAHD 58
Db 154 SNGQPGIGPPGPRGSGEYVPGPPGPPGPPGPGIDMSAPAGLSQP--EKGPD 211
Qy 59 GGRYRADDANVVRDRDLEVDITLKSLSQIENIRSPGSRKNPARTCDLKNCHSDWK 117
Db 212 PLRYMRADREASSLRQHDVEVDATLKSINGQIEDIRSPGSRKNPARSBDLKLCHPEWK 271
Qy 118 SGSEYIWDPNQCNLDALKVCNMTGTCVVPQPSVAQKNWYISKNPDKRHVWFQESM 177
Db 272 SGDYWDVDFNLGSAADALKVCNMTGTCVVPSTPKIPRKNWYTSKS-KAQKHVWFQESM 330
Qy 178 TDGFQFFYGGQSDPADVAIQTLFLRLMSTEAQNITYHCNKSVAYMDQGTGKLLKALL 237
Db 331 NGGFHFSYADGSGTPTTTIQLNPLRLSTEAQTITYHCNKSVAYMDQATGKLLKALL 390
Qy 238 KGSNEIRAEGRSFRFYVTVDCGTSHTGAWGKTVIEYKTKSSRLPIIDVAPLDVGAP 297
Db 391 QGSNDVIRAEGRSFRFYGVLEDCGCKKTKQWAKTVIEYKTKTSRLPIINDIAPMDIGGA 450
Qy 298 DQFEGFDVGPFVCL 311
Db 451 DQFEGVDIGAVCL 464

RESULT 9
B40333
collagen alpha 1(II) chain precursor - African clawed frog
C/Species: Xenopus laevis (African clawed frog)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: B40333
R/Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, P.
J. Cell Biol. 115, 565-575, 1991
A/Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis em
A/Reference number: A40333; MUID:92011898; PMID:1918153
A/Accession: B40333
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-1486 <SUA>
A/Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; UNIPARC:UPI0000173B50; GB:M63595
C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
C/Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix
F/37-96/Domain: von Willebrand factor type C repeat homology <VMC>
F/1256-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 68.8%; Score 1175; DB 1; Length 1486;
Best Local Similarity 66.1%; Pred. No. 7.8e-79;
Matches 207; Conservative 45; Mismatches 53; Indels 8; Gaps 4;

Qy 2 SNGLPFGIPGPPGRGTGAGVPGPPGPPGPPSAGFD---FSFLPQPPEKAHD 58
Db 1179 SNGISGIPGPPGPRGSGEYVPGPPGPPGPPGPGIDMSAPAGLSQP--EKGPD 1236
Qy 59 GGRYRADDANVVRDRDLEVDITLKSLSQIENIRSPGSRKNPARTCDLKNCHSDWKS 118

Db	1237	PMYMRADQAS--NSLPVDVEATLKSUNNIGENIRSPDGTGKOPARTCDRLKI	CHPSEWKS	1294
Qy	119	GXYWIDPNQCNDLAIKVFCNNMETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMT	178	
Db	1295	GDYWIDPNQCETDAIKVFCDMETGETCVYPNSKIPKKNWSAKG-KEKKHIWFGETIN	1353	
Qy	179	DGFQFEGYGGSDDPADVAIQLTFLRLMSTEASONITYHCNKNVAYMQOQTGNLKALLK	238	
Db	1354	GGFQFSYGDSSAPNTANIQMTFLRLISTDASQNITYHCNKNIAFMDEASGNLKAVALQ	1413	
Qy	239	GSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVLEYKTKTSRLPIIDVAPLDVGAP	298	
Db	1414	GSNDVEIRAEGNSRFTYNALDGCKHGTGWSKTVIEYRTQKTSRLPIVIDIAPMDIGAD	1473	
Qy	299	QEFGFDVGPVCFL	311	
Db	1474	QEFGVDIGPVCFL	1486	
 RESULT 10 A40333 collagen alpha 1'(II) chain precursor - African clawed frog C:Species: Xenopus laevis (African clawed frog) C>Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 09-Jul-2004 C/Accession: A40333 R/Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F. J. Cell Biol. 115, 565-575, 1991 A>Title: Expression of two nonallelic type II procollagen genes during Xenopus A/Reference number: A40333; PMID:92011898; PMID:1918153 A/Accession: A40333 A>Status: nucleic acid sequence not shown A/Molecule type: mRNA A/Residues: 1-1492 <SUA> A/Cross-references: UNIPROT:Q91718; UNIPROT:Q91717; UNIPARC:UIPO000177382; GB:M-M- es they replace; the appropriate interpretation of the sequence figure was reco C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal C/Keywords: coiled coil, extracellular matrix; glycoprotein; trimer; triple hel P/37-96/Domain: von Willebrand factor type C repeat homology <VMC> P/1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>				
 Query Match 67.9%; Score 1159.5; DB 2; Length 1492; Best Local Similarity 65.3%; Pred. No. 1.1e-77; Matches 205; Conservative 47; Mismatches 55; Indels 7; Gaps 4;				
Qy	2	SNGLGPIGPGPRGRGTAGPVGPPGPPGPPGPPGPSAGFD---PSFLPQPPEKAHD	58	
Db	1182	SNGIPGPIGPPGRSGTGPAGPGSQGPPGPPGPPGPDMSAFAGLSQP--EKGPD	1239	
Qy	59	GGRYRADADANVRDLRDLEDTTLKSLSQIEN-IRSPEGSRKNPARTCDLKMCHSDWK	117	
Db	1240	PMYMRADQASSVLQTDADVANKLTINNQIESIIISPDGTKKNPARTCDLKLCHPEWK	1299	
Qy	118	SGBYWIDPNQCNLDIAIKVFECNMETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESM	177	
Db	1300	SGEYWI DPNEGCTVDAIKVFECNMETGESCVYPSKIPKKNWSAKG-KEKKHIWFGETI	1358	
Qy	178	TDFQFEGYGGSDDPADVAIQLTFLRLMSTEASONITYHCNKNVAYMQOQTGNLKALLK	237	
Db	1359	NGGFQFSYGDSSAPNTANIQMTFLRLISTDASQNITYHCNKNIAFMDEASGNLKAVALL	1418	
Qy	238	KGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVLEYKTKTSRLPIIDVAPLDVGAP	297	
Db	1419	QSNDVEIRAEGNSRFTYNALDGCKHGTGWSKTVIEYRTQKTSRLPIVIDIAPMDIGGA	1478	
Qy	298	DQFEGFDVGPVCFL	311	
Db	1479	DQFEGVDIGPVCFL	1492	
 RESULT 11 CGGH6C				

A:Molecule type: DNA
A:Residues: 536-558 <AVV>
A:Cross-references: UNIPARC:UPI00001712AD; GB:M10581; NID:G211323; PIDN:AAA48637.1; PID:
R:Lehrach, H.; Frischach, A.M.; Hanahan, D.; Wozney, J.; Fuller, F.; Crkvenjakov, R.; Bo.
Proc. Natl. Acad. Sci. U.S.A. 75, 5417-5421, 1978
A>Title: Construction and characterization of a 2.5-kilobase procollagen clone.
A:Reference number: I50171; MUID:79074829; PMID:364479
A:Accession: I50171
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 484-505 <LEH>
A:Cross-references: UNIPARC:UPI00001712AB; GB:J00837; NID:G4530617; PIDN:AAA51614.1; PID
C:Genetics:
A:Gene: COL1A2
A:Introns: 24/1; 27/3; 33/3; 45/3; 73/3; 92/3; 107/3; 125/3; 143/3; 161/3; 179/3; 197/3;
A>Note: the list of introns is incomplete
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C:Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxypro
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-77/Domain: amino-terminal propeptide #status predicted <PRO>
F:78-964/Product: collagen alpha 2(I) chain (fragments) #status predicted <MATN>
F:78-89/Region: amino-terminal nonhelical telopeptide
F:737-964/Domain: fibrillar collagen carboxyl-terminal homology <FC>
F:78/Modified site: blocked amino end (Gln) (in mature form) (probably pyrrolidone carb
F:83/Modified site: alanine (Lys) #status experimental
F:422,425/Modified site: 4-hydroxyproline (Pro) #status predicted
F:866/Binding site: carbohydrate (Asn) (covalent) #status predicted

	Query Match	Best Local Similarity	Score	Length
	Matches 185;	Conservative 46;	Mismatches 63;	Indels 14; Gaps 4;

Qy 3 NGLPGPIGPGRGTGDAGVPGGPPGPDPGPSAGDFSLPQPPOEKAHDGGRY 62
Db 670 NGLFGPIGPAGVRSGSHSQGAPGPDPGPDPGGGGEVGFP-----DAEY 718

Qy 63 YRADANVVRRDLVDPTTLKLSQQIENIRSPESGRKNPARTCDRLKMCHSDWKSGEYW 122
Db 719 YRAQPS-LRPKDVEVDATLTNNQIETLLTPEGSKKNPARTCDRLRLSHPEWSGFGYW 777

Qy 123 IDPNQGNCILDAIKVFCEMETGETCYVTPQPSVAQNWIYSKNPKDKRHWFGESEMTDGFQ 182
Db 778 IDPNQGGCTADAIRAYCDPATGETCIHASLEDIPTKTYYVSKNPKDKKHFWFGE-INGGTQ 836

Qy 183 FEYGCGSDPADVAIQLTFLRLMSTEAQNITYHCNSVAYNDQQTGNLKCALLLGKSN 242
Db 837 FEYNAGEGYTKDMATQLAFMLLANHASQNTIYHCNSIAYMDEETGNLKKXAVILOGSD 896

Qy 243 IEIRAEGNSREFTSYVTVDGCTSHTGAWKTIIEYKTTKSSRLPIIDVAPLDVGAPQDEF 302
Db 897 VELRAEG-SRFTFSVLVDGCCKNKMKWKTIIIEYRTNKPRLPILDIAPLDIGGAOEF 955

Qy 303 FDVGPFVCF 310
Db 956 LHIGFVCF 963

RESULT 13
S10847
collagen alpha 2(I) chain - chicken (fragment)
C:Species: Gallus gallus (chicken)
C>Date: 21-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 23-May-1997
C:Accession: S10847; S65690
E:Dickson, L.A.; Ninomiya, Y.; Bernard, M.P.; Pesciotta, D.M.; Parsons, J.; Green, G.; E
J. Biol. Chem. 256, 8407-8415, 1981
A>Title: The exon/intron structure of the 3'-region of the pro-alpha-2(I) collagen gene.
A:Reference number: S10847; MUID:81264246; PMID:6267043
A:Accession: S10847
A:Molecule type: DNA
A:Residues: 1-365 <DIC>
A:Cross-references: UNIPARC:UPI0000177396; EMBL:J00811
A>Note: the authors translated the codon ATA for residue 207 as Asp, AGC for residue 216
A:Accession: S65690

A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm	Db	1450	GGPDQBFQFVDVGPVCF 1466
A:Reference number: 159025; MUID:85216505; PMID:3858826			
A:Accession: I79359			
A:Status: translated from GB/EMBL/DBJ			
A:Molecule type: mRNA			
A:Residues: 1165-1196 <MA>			
A:Cross-references: UNIPARC:UPI000016A6B6; GB:M11134; NID:g180417; PIDN:AAA52004.1; PID:			
R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.			
J. Biol. Chem. 260, 4357-4363, 1985			
A:Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. F			
A:Reference number: A92516; MUID:85157600; PMID:2579949			
A:Accession: A92516			
A:Molecule type: DNA			
A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>			
A:Cross-references: UNIPARC:UPI0000173B89; GB:M10615; GB:M10793; GB:M10794; GB:M10795; G			
A:Experimental source: liver			
A:Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given B			
ation			
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit (
3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently C			
C:Genetics:			
A:Gene: GDB:COL3A1			
A:Cross-references: GDB:118729; OMIM:120180			
A:Map position: 2q31-2q31			
A:Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3			
A:Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan			
C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b			
er of their length, is formed with desmosine cross-links made from lysine and allylsine r			
C:Function:			
A:Description: structural component of extracellular fibrous polymer that maintains inte			
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; h			
C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd			
F:1-23/Domain: signal sequence #status predicted <SIG>			
F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>			
F:31-91/Domain: von Willebrand factor type C repeat homology <VMC>			
F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>			
F:154-167/Region: amino-terminal nonhelical telopeptide			
F:168-1196/Region: helical			
F:1091-1093/Region: cell attachment (R-G-D) motif			
F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide			
F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>			
F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>			
F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted			
F:153-154/Cleavage site: Pro-Gln (procollagen N-endorpeptidase) #status predicted			
F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predi			
F:161,1212/Modified site: allylsine (Lys) #status predicted			
Query Match	59.0%;	Score 1007;	DB 1; Length 1466;
Best Local Similarity	60.3%;	Pred. No. 2.1e-66;	
Matches 191; Conservative 39; Mismatches 77; Indels 10; Gaps 5;			
Qy	2	SNGLPGTIGPPGRGTGAGVGP	PPGPPGPPGPPSA-----GFDFSLFPQPQEK 55
Db	1153	TSGHPTIGPPGRNGRSESG	PGHPGPPGPPGPPGCGGVAIAIGIGEX 1212
Qy	56	AHDGRRYRADDANVVRDRD	LEVDTTLKLSQQIENRISPEGRNRPARTCRDLKMCSD 115
Db	1213	AGGFAPYGGDEPMDFKINTD	-EIMTSLKSANGQIESLISPDGRKNPARNCRDLKFCHPE 1271
Qy	116	WKSGEYWDNOCNLDIAIKV	FCNMETGTCVPTQPSVAGKNWYISKNPDKKHVWFGE 175
Db	1272	LKSGEYWDPNQCGKLDIAIK	VFCNMETGTCISANPLNVPKHWV-TDSSAEKKHVVWFGE 1330
Qy	176	SMTDGFQFEGCGSDPADVA	-IQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKK 234
Db	1331	SMDCGFQFSGNYP-ELPED	VDLDVQLAFRLLSRSASQNIYHCKNSIAYMDQASGVKKA 1389
Qy	235	LLKGSNEIEIRAGNSRFTY	SVTVGCTSHGAWKTVIEYKTKSRLPIDVAPLDV 294
Db	1390	LKMGSEGEFKAEGNSKFTY	LVLEDGCTKHTGEWSKTVFEYRTRKAVRLPIVDIAPYDI 1449
Qy	295	GAPDQBFQFVDVGPVCF 311	

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:13:00 ; Search time 169.434 Seconds
(without alignments)
1295.016 Million cell updates/sec

Title: US-10-677-877A-2

Perfect score: 1707

Sequence: 1 RSNGLPICPPGPRGTG.....LDVGAPDQFDPVGVCF 311

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1695	99.3	1461	2	O76045 homo sapien
2	1695	99.3	1464	1	COL1A1_HUMAN
3	1691	99.1	1467	2	O59F64 HUMAN
4	1673	98.0	1464	2	O8N473 homo sapien
5	1664	97.5	1460	1	COL1A1_CANFA
6	1591	93.2	1453	1	COL1A1_CHICK
7	1591	93.2	1453	2	Q63079 rattus norv
8	1588	93.0	589	2	Q99LL6 MOUSE
9	1588	93.0	1225	2	Q6PCL3 MOUSE
10	1588	93.0	1453	2	Q81039 MOUSE
11	1564	91.6	1453	1	COL1A1_MOUSE
12	1543	90.4	325	2	Q60785 MOUSE
13	1534	89.9	1450	2	Q9Y1B4_CYNPY
14	1528	89.5	287	2	Q95ND8 BOVIN
15	1506.5	88.3	1449	2	Q64082 xenopus tro
16	1490	87.3	1445	2	Q93251 RANCA
17	1483.5	86.9	1449	2	Q802B5 xenopus lae
18	1448	84.8	1447	2	Q9IB91 xenopus lae
19	1427	83.6	809	2	Q93485 oncorhynch
20	1427	83.6	1449	2	Q910C0 ONCMY
21	1403.5	82.2	481	2	Q90YJ1 brachydanio
22	1403.5	82.2	1447	2	Q6P4U1 BRARE
23	1403.5	82.2	1447	2	Q6ULJ5 BRARE
24	1372.5	80.4	1447	2	Q5NT96 PAROL
25	1330	77.9	1463	2	Q4W6W6 9CHON
26	1313	76.9	1545	2	Q4RYF7 TETNG
27	1274	74.6	1449	2	Q6NZ15 BRARE
28	1274	74.6	1449	2	Q6PE19 BRARE
29	1266.5	74.2	678	2	Q93486 ONCMY
30	1266.5	74.2	1458	2	Q910B9 ONCMY
31	1262.5	74.0	441	2	Q90Y19 brachydanio

32 1250.5 73.3 1399 2 O4SB89 TETNG
33 1236.5 72.4 369 1 CO2A1_CHICK
34 1229.5 72.0 1269 2 Q7T2Z7 CHICK
35 1223.5 71.7 1420 2 Q90W37 gallus gall
36 1220.5 71.5 1418 2 Q9W7R9 CYNPY
37 1213.5 71.1 1418 1 CO2A1_HUMAN
38 1213.5 71.1 1487 2 O14047 homo sapien
39 1213.5 71.1 1487 2 O77753 canis fami
40 1212.5 71.0 1418 2 Q28396 equus cabal
41 1206.5 70.7 826 2 Q8K0N6 mus musculu
42 1206.5 70.7 886 2 Q8CEF7 MOUSE
43 1206.5 70.7 1419 2 Q80VY3 mus musculu
44 1206.5 70.7 1419 2 Q80X38 mus musculu
45 1206.5 70.7 1419 2 Q63123 RAT

ALIGNMENTS

RESULT 1

O76045 HUMAN
ID O76045 HUMAN PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
RA Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1).";
RL Gene 67:105-115(1988).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockup D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
RT COL1A1 gene bind cell-specific nuclear proteins.";
RL FEBS Lett. 279:9-13(1991).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
RA Olsen A.S., Prockup D.J.;
RT "Completion of the last half of the structure of the human gene for
RT the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
RL Matrix 11:375-379(1991).
RN [6]

Search completed: April 26, 2006, 16:19:22
Job time : 172.434 secs

Db 1393 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTTKTSRLPIIDVAPLDVGAPDQFEG 1452
Qy 303 FDVGPPVCF 311
Db 1453 FDVGPPVCF 1461

RESULT 2
US-09-289-578-9
; Sequence 9, Application US/09289578
; Patent No. 6428978
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; CURRENT FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

Query Match 99.3%; Score 1695; DB 2; Length 1461;
Best Local Similarity 99.7%; Pred. No. 1.6e-141;
Matches 308; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 3 NGLPGPIGPPGRGRTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPQEKADGGRY 62
Db 1153 NGLPGPIGPPGRGRTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPQEKADGGRY 1212
Qy 63 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMCSDWKSGEYW 122
Db 1213 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMCSDWKSGEYW 1272
Qy 123 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQ 182
Db 1273 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQ 1332
Qy 183 FEYGGGSDPADVAIQLTFLRLMSTASQNIYHCKNSVAYMDQOTGNLKKALLKGSNE 242
Db 1333 FEYGGGSDPADVAIQLTFLRLMSTASQNIYHCKNSVAYMDQOTGNLKKALLKGSNE 1392
Qy 243 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTTKTSRLPIIDVAPLDVGAPDQFEG 302
Db 1393 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTTKTSRLPIIDVAPLDVGAPDQFEG 1452
Qy 303 FDVGPPVCF 311
Db 1453 FDVGPPVCF 1461

RESULT 3
US-09-331-347C-21
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
; TITLE OF INVENTION: Obtaining Such and Their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331,347C

; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-331-347C-21
Query Match 99.3%; Score 1695; DB 2; Length 1464;
Best Local Similarity 99.7%; Pred. No. 1.6e-141;
Matches 308; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 3 NGLPGPIGPPGRGRTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPQEKADGGRY 62
Db 1156 NGLPGPIGPPGRGRTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPQEKADGGRY 1215
Qy 63 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMCSDWKSGEYW 122
Db 1216 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMCSDWKSGEYW 1275
Qy 123 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQ 182
Db 1276 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQ 1335
Qy 183 FEYGGGSDPADVAIQLTFLRLMSTASQNIYHCKNSVAYMDQOTGNLKKALLKGSNE 242
Db 1336 FEYGGGSDPADVAIQLTFLRLMSTASQNIYHCKNSVAYMDQOTGNLKKALLKGSNE 1395
Qy 243 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTTKTSRLPIIDVAPLDVGAPDQFEG 302
Db 1396 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTTKTSRLPIIDVAPLDVGAPDQFEG 1455
Qy 303 FDVGPPVCF 311
Db 1456 FDVGPPVCF 1464
RESULT 4
US-08-963-825-18
; Sequence 18, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/570,573
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/187,319
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
US-09-570-573-18
Query Match 97.9%; Score 1670.5; DB 2; Length 1341;
Best Local Similarity 98.7%; Pred. No. 2.1e-139;
Matches 306; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 3 NGLPGIGPPGPRGRTGDAGPVGPPGPPGPPSAGDFSFLLPQPPQEKAHKGGRY 62
DB 1032 NGLPGIGPPGPRGRTGDAGPVGPPGPPGPPSAGDFSFLLPQPPQEKAHKGGRY 1091
QY 63 YRADDANVVRDRLEVDVTTLSQIENIRSPGSRKNPARTCRDLKCHSDWKSGEY 122
DB 1092 YRADDANVVRDRLEVDVTTLSQIENIRSPGSRKNPARTCRDLKCHSDWKSGEY 1151
QY 123 IDPNQCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDK-RHVVFGESMTDGF 181
DB 1152 IDPNQCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHRVWFGESMTDGF 1211
QY 182 QFEYGGGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSN 241
DB 1212 QFEYGGGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSN 1271
QY 242 EIEIRAGNSRFTYSVTVDGCTSHTGAWGKTVEYTKTKSSRLPIIDVAPLDVGAPDQEF 301
DB 1272 EIEIRAGNSRFTYSVTVDGCTSHTGAWGKTVEYTKTKSSRLPIIDVAPLDVGAPDQEF 1331
QY 302 GFDVGPVCFL 311
DB 1332 GFDVGPVCFL 1341
RESULT 7
US-09-548-608-18
Sequence 18, Application US/09548608
Patent No. 6355442
GENERAL INFORMATION:
APPLICANT: Qvist, Per
APPLICANT: Bonde, Martin
TITLE OF INVENTION: A Method for Assaying Collagen Fragments
TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
TITLE OF INVENTION: Disorders Associated with the Metabolism of
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/548,608
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/187,319
APPLICATION NUMBER: 08/187,319
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gogoris, Adda C
REGISTRATION NUMBER: 29,714
REFERENCE/DOCKET NUMBER: 4305/08701
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-527-7700
TELEFAX: 212-753-6237
TELEX: 236687
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1341 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: COLLAGEN ALPHA 1 (I)
US-09-548-608-18
Query Match 97.9%; Score 1670.5; DB 2; Length 1341;
Best Local Similarity 98.7%; Pred. No. 2.1e-139;
Matches 306; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 3 NGLPGIGPPGPRGRTGDAGPVGPPGPPGPPSAGDFSFLLPQPPQEKAHKGGRY 62
DB 1032 NGLPGIGPPGPRGRTGDAGPVGPPGPPGPPSAGDFSFLLPQPPQEKAHKGGRY 1091
QY 63 YRADDANVVRDRLEVDVTTLSQIENIRSPGSRKNPARTCRDLKCHSDWKSGEY 122
DB 1092 YRADDANVVRDRLEVDVTTLSQIENIRSPGSRKNPARTCRDLKCHSDWKSGEY 1151
QY 123 IDPNQCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDK-RHVVFGESMTDGF 181
DB 1152 IDPNQCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHRVWFGESMTDGF 1211
QY 182 QFEYGGGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSN 241
DB 1212 QFEYGGGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSN 1271
QY 242 EIEIRAGNSRFTYSVTVDGCTSHTGAWGKTVEYTKTKSSRLPIIDVAPLDVGAPDQEF 301
DB 1272 EIEIRAGNSRFTYSVTVDGCTSHTGAWGKTVEYTKTKSSRLPIIDVAPLDVGAPDQEF 1331
QY 302 GFDVGPVCFL 311
DB 1332 GFDVGPVCFL 1341
RESULT 8
US-09-029-348-14
Sequence 14, Application US/09029348
Patent No. 6171827
GENERAL INFORMATION:
APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
TITLE OF INVENTION: NOVEL PROCOLLAGENS

Db 1289 HPSYDDNLA PNTANVQMTFLRLLS TEGSQNIYHCKNCIAYLDEAAGNLKKALLIQSN 1348
Qy 242 EIEIRAGNSRFTYSVTVGCTSHTGAWGKTVEYKTKSRLPFIIDVAPLDVGAPDOEF 301
Db 1349 DVEIRAGNSRFTYTALKDGCCTKHTGKWKTVIEYRSQKTSRLPFIIDAPMDIGGPQEF 1408
Qy 302 GPDVGPVCFEL 311
Db 1409 GVDIGPVCFEL 1418

Search completed: April 26, 2006, 16:18:00
Job time : 24.4086 secs

THIS PAGE LEFT BLANK

[illegible][illegible]

```

, PRIOR FILING DATE: 2001-06-13
, PRIOR APPLICATION NUMBER: US 60/298,155
, PRIOR FILING DATE: 2001-06-13
, PRIOR APPLICATION NUMBER: US 60/335,936
, PRIOR FILING DATE: 2001-11-14
, NUMBER OF SEQ ID NOS: 238
, SOFTWARE: Fast-Seq for Windows Version 4.0
, SEQ ID NO 36
, LENGTH: 1464
, TYPE: PRT
, ORGANISM: Homo sapiens
US-10-171-311-36

Query Match          99.5%; Score 1698; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.8e-129;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3  NGLPGPIGPPGPRGTGDAGVGGPPGPPGPPGPPSAGFDPSFLPQPPOEKAHDGGRY  62
      |||
Db      1156 NGLPGPIGPPGPRGTGDAGVGGPPGPPGPPGPPSAGFDPSFLPQPPOEKAHDGGRY  1215
      |||

Qy      63  YRADDANVVRDRDLEVDVTTLKLSQQTENIRSPGSRKKNPAPTRCDLKMCHSDWKSGEYW  122
      |||
Db      1216 YRADDANVVRDRDLEVDVTTLKLSQQTENIRSPGSRKKNPAPTRCDLKMCHSDWKSGEYW  1275
      |||

Qy      123 IDPNQGCNLDIAKVFCCNMETGETCVYPTOPSAVQKNWYISKNPDKRHHVWFSGSMTDGFQ  182

```

Db 1276 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 1335
Qy 183 FEYGGQGSDDADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNE 242
Db 1336 FEYGGQGSDDADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNE 1395
Qy 243 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ 302
Db 1396 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ 1455
Qy 303 FDVGPVCFL 311
Db 1456 FDVGPVCFL 1464

RESULT 8
US-10-149-352-2
; Sequence 2, Application US/10149352
; Publication No. US20030105050A1
; GENERAL INFORMATION:
; APPLICANT: Berl, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/10/149,352
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-149-352-2

Query Match 99.5%; Score 1698; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.8e-129;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 NGLPGPIGPPGRGRTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPPQKHAHDGGRY 62
Db 1156 NGLPGPIGPPGRGRTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPPQKHAHDGGRY 1215
Qy 63 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 122
Db 1216 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 1275
Qy 123 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 182
Db 1276 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 1335
Qy 183 FEYGGQGSDDADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNE 242
Db 1336 FEYGGQGSDDADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNE 1395
Qy 243 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ 302
Db 1396 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ 1455
Qy 303 FDVGPVCFL 311
Db 1456 FDVGPVCFL 1464

RESULT 9
US-10-177-293-65
; Sequence 65, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James

; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John E.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: East Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; FILE REFERENCE: MRI-038
; FILE REFERENCE: PREVENTION, AND THERAPY OF BREAST CANCER
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 65
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-65

Query Match 99.5%; Score 1698; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.8e-129;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 NGLPGPIGPPGRGRTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPPQKHAHDGGRY 62
Db 1156 NGLPGPIGPPGRGRTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPPQKHAHDGGRY 1315
Qy 63 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 122
Db 1216 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 1275
Qy 123 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 182
Db 1276 IDNQCGLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 1335
Qy 183 FEYGGQGSDDADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNE 242
Db 1336 FEYGGQGSDDADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNE 1395
Qy 243 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ 302
Db 1396 IEIRAEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ 1455
Qy 303 FDVGPVCFL 311
Db 1456 FDVGPVCFL 1464
RESULT 10
US-10-301-822-28

```
; Sequence 28, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/10/301,822
; CURRENT FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-301-822-28

Query Match          99.5%; Score 1698; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.8e-129;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  3  NGLPGPIGPPGRGTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPPOEKAHDGGRY  62
Db  1156 NGLPGPIGPPGRGTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPPOEKAHDGGRY  1215

Qy  63  YRADDANVVRDRDLEVDVTLKSLSQOIENTRSPGSRKNPARTCDLKMCHSDWKSGEYW  122
Db  1216 YRADDANVVRDRDLEVDVTLKSLSQOIENTRSPGSRKNPARTCDLKMCHSDWKSGEYW  1275

Qy  123 IDPNOGNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKRVHVFGESEMTDGFQ  182
Db  1276 IDPNOGNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKRVHVFGESEMTDGFQ  1335

Qy  183 FEYGGQSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNE  242
Db  1336 FEYGGQSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNE  1395

Qy  243 IEIRAEGNSRFTYSVTVDGCTSHTGAMGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ  302
Db  1396 IEIRAEGNSRFTYSVTVDGCTSHTGAMGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ  1455

Qy  303 FDVGPVCFL 311
Db  1456 FDVGPVCFL 1464

RESULT 11
US-10-734-564-79
; Sequence 79, Application US/10734564
; Publication No. US20040157278A1
; GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using TIMP1
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10/734,564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
```

```
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-734-564-79

Query Match          99.5%; Score 1698; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.8e-129;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  3  NGLPGPIGPPGRGTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPPOEKAHDGGRY  62
Db  1156 NGLPGPIGPPGRGTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPPOEKAHDGGRY  1215

Qy  63  YRADDANVVRDRDLEVDVTLKSLSQOIENTRSPGSRKNPARTCDLKMCHSDWKSGEYW  122
Db  1216 YRADDANVVRDRDLEVDVTLKSLSQOIENTRSPGSRKNPARTCDLKMCHSDWKSGEYW  1275

Qy  123 IDPNOGNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKRVHVFGESEMTDGFQ  182
Db  1276 IDPNOGNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKRVHVFGESEMTDGFQ  1335

Qy  183 FEYGGQSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNE  242
Db  1336 FEYGGQSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNE  1395

Qy  243 IEIRAEGNSRFTYSVTVDGCTSHTGAMGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ  302
Db  1396 IEIRAEGNSRFTYSVTVDGCTSHTGAMGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ  1455

Qy  303 FDVGPVCFL 311
Db  1456 FDVGPVCFL 1464

RESULT 12
US-10-474-794-261
; Sequence 261, Application US/10474794
; Publication No. US20040213793A1
; GENERAL INFORMATION:
; APPLICANT: Carson-Walter, Eleanor
; APPLICANT: St. Croix, Brad
; APPLICANT: Vogelstein, Bert
; APPLICANT: Kinzler, Kenneth
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00179
; CURRENT APPLICATION NUMBER: US/10/474,794
; CURRENT FILING DATE: 2003-10-14
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 60/308,829
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 359
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 261
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-474-794-261

Query Match          99.5%; Score 1698; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.8e-129;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy  3  NGLPGPIGPPGRGTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPPOEKAHDGGRY  62
Db  1156 NGLPGPIGPPGRGTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPPOEKAHDGGRY  1215

Qy  63  YRADDANVVRDRDLEVDVTLKSLSQOIENTRSPGSRKNPARTCDLKMCHSDWKSGEYW  122
Db  1216 YRADDANVVRDRDLEVDVTLKSLSQOIENTRSPGSRKNPARTCDLKMCHSDWKSGEYW  1275

Qy  123 IDPNOGNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKRVHVFGESEMTDGFQ  182
Db  1396 IDPNOGNCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKRVHVFGESEMTDGFQ  1455
```

Db 1276 IDPNOGCLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 1335
Qy 183 FEYGGGSDPADVAIQLTFLRLMSTASQNIITHCKNSVAYMDQQTGNLKKALLKGSNE 242
Db 1336 FEYGGGSDPADVAIQLTFLRLMSTASQNIITHCKNSVAYMDQQTGNLKKALLKGSNE 1395
Qy 243 IEIRAEGNSRFTYSVTVGCTSHTGAGTKTVIEYKTKSSRLPIIDVAPLDVGAPDQEF 302
Db 1396 IEIRAEGNSRFTYSVTVGCTSHTGAGTKTVIEYKTKSSRLPIIDVAPLDVGAPDQEF 1455
Qy 303 FDVGPVCFL 311
Db 1456 FDVGPVCFL 1464

RESULT 13
US-10-723-860-2289
; Sequence 2289, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882, 0193, NPUS01
; CURRENT APPLICATION NUMBER: US/10723,860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429,739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2289
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723-860-2289

Query Match 99.5%; Score 1698; DB 5; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.8e-129;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 NGLPGPIGPPGRGTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPPQEKADHGGRY 62
Db 1156 NGLPGPIGPPGRGTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPPQEKADHGGRY 1215
Qy 63 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 122
Db 1216 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 1275
Qy 123 IDPNOGCLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 182
Db 1276 IDPNOGCLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 1335
Qy 183 FEYGGGSDPADVAIQLTFLRLMSTASQNIITHCKNSVAYMDQQTGNLKKALLKGSNE 242
Db 1336 FEYGGGSDPADVAIQLTFLRLMSTASQNIITHCKNSVAYMDQQTGNLKKALLKGSNE 1395
Qy 243 IEIRAEGNSRFTYSVTVGCTSHTGAGTKTVIEYKTKSSRLPIIDVAPLDVGAPDQEF 302
Db 1396 IEIRAEGNSRFTYSVTVGCTSHTGAGTKTVIEYKTKSSRLPIIDVAPLDVGAPDQEF 1455
Qy 303 FDVGPVCFL 311
Db 1456 FDVGPVCFL 1464

RESULT 14
US-10-852-335A-157
; Sequence 157, Application US/10852335A
; Publication No. US20050112129A1
; GENERAL INFORMATION:
; APPLICANT: HEIDI S. PHILLIPS

; TITLE OF INVENTION: Compositions and Methods for the Diagnosis and
; TITLE OF INVENTION: Treatment of Tumors of Glial Origin
; FILE REFERENCE: P5103R1-US
; CURRENT APPLICATION NUMBER: US/10/852,335A
; CURRENT FILING DATE: 2004-05-24
; PRIOR APPLICATION NUMBER: US 60/548,299
; PRIOR FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/473,238
; PRIOR FILING DATE: 2003-05-23
; NUMBER OF SEQ ID NOS: 190
; SEQ ID NO 157
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-852-335A-157

Query Match 99.5%; Score 1698; DB 5; Length 1464;
Best Local Similarity 100.0%; Pred. No. 5.8e-129;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 NGLPGPIGPPGRGTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPPQEKADHGGRY 62
Db 1156 NGLPGPIGPPGRGTGDAGVPVGGPPGPPGPPSAGDFSFLLPQPPQEKADHGGRY 1215
Qy 63 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 122
Db 1216 YRADDANVVRDRDLEVDVTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 1275
Qy 123 IDPNOGCLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 182
Db 1276 IDPNOGCLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 1335
Qy 183 FEYGGGSDPADVAIQLTFLRLMSTASQNIITHCKNSVAYMDQQTGNLKKALLKGSNE 242
Db 1336 FEYGGGSDPADVAIQLTFLRLMSTASQNIITHCKNSVAYMDQQTGNLKKALLKGSNE 1395
Qy 243 IEIRAEGNSRFTYSVTVGCTSHTGAGTKTVIEYKTKSSRLPIIDVAPLDVGAPDQEF 302
Db 1396 IEIRAEGNSRFTYSVTVGCTSHTGAGTKTVIEYKTKSSRLPIIDVAPLDVGAPDQEF 1455
Qy 303 FDVGPVCFL 311
Db 1456 FDVGPVCFL 1464

RESULT 15
US-10-979-159-261
; Sequence 261, Application US/10979159
; Publication No. US20050142138A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOGENOUS CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107, 00134
; CURRENT APPLICATION NUMBER: US/10/979,159
; CURRENT FILING DATE: 2004-11-03
; PRIOR APPLICATION NUMBER: US/09/918,715
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 261
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-979-159-261

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:18:13 ; Search time 12.2616 Seconds
(without alignments)
1153.484 Million cell updates/sec

Title: US-10-677-877a-2
Perfect score: 1707
Sequence: 1 RSNGLPPTGPPGPRGTGD.....LDVGAPDQDFGVPCFL 311

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 232119 seqs, 45477862 residues

Total number of hits satisfying chosen parameters: 232119

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.*
1: /SID55/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /SID55/ptodata/2/pubpaa/US06_NEW_PUB pep.*
3: /SID55/ptodata/2/pubpaa/US07_NEW_PUB pep.*
4: /SID55/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
5: /SID55/ptodata/2/pubpaa/US05_NEW_PUB pep.*
6: /SID55/ptodata/2/pubpaa/US10_NEW_PUB pep.*
7: /SID55/ptodata/2/pubpaa/US11_NEW_PUB pep.*
8: /SID55/ptodata/2/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1698	99.5	1464	6 US-10-501-035-331	Sequence 331, App
2	1698	99.5	1464	7 US-11-186-284-28	Sequence 28, Appl
3	1698	99.5	1464	7 US-11-021-603-2	Sequence 2, Appli
4	1691	99.1	1464	7 US-11-000-263-243	Sequence 243, App
5	1691	99.1	1467	6 US-10-821-234-1096	Sequence 1096, Ap
6	1591	93.2	1453	6 US-10-784-004-344	Sequence 344, App
7	1591	93.2	1453	6 US-10-784-004-434	Sequence 434, App
8	1591	93.2	1453	6 US-10-784-004-918	Sequence 918, App
9	1591	93.2	1453	6 US-10-784-004-950	Sequence 950, App
10	1213.5	71.1	1487	7 US-11-202-057-3	Sequence 3, Appli
11	1213.5	71.1	1487	7 US-11-202-057-5	Sequence 5, Appli
12	1213.5	71.1	1487	7 US-11-202-057-7	Sequence 7, Appli
13	1024	60.0	1366	6 US-10-821-234-1431	Sequence 1431, Ap
14	1024	60.0	1366	7 US-11-186-284-31	Sequence 31, Appl
15	1009	59.1	1166	6 US-10-821-234-964	Sequence 964, App
16	1009	59.1	1466	7 US-11-186-284-33	Sequence 33, Appl
17	1003	58.8	636	6 US-10-784-004-428	Sequence 428, App
18	912.5	53.5	218	7 US-11-202-057-28	Sequence 28, Appl
19	909.5	53.3	1496	7 US-11-186-284-35	Sequence 35, Appl
20	640	37.5	1876	6 US-10-821-234-1182	Sequence 1182, Ap
21	639	37.4	1806	7 US-11-051-720-1446	Sequence 1446, Ap
22	639	37.4	1806	7 US-11-051-720-1447	Sequence 1447, Ap
23	633	37.1	1767	6 US-10-995-561-911	Sequence 911, App
24	633	37.1	1767	6 US-10-995-561-914	Sequence 914, App
25	633	37.1	1806	6 US-10-995-561-912	Sequence 912, App

RESULT 1

US-10-501-035-331
; Sequence 331, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501,035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 331
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-501-035-331

Query Match 99.5%; Score 1698; DB 6; Length 1464;
Best Local Similarity 100.0%; Pred. No. 2.7e-132;
Matches 309; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	3	NGLPGPPIGPPRGRTGDAGVGP	PPGPPGPPPSAGDFSF	LPQPPQEKADGGRY	62
Db	1156	NGLPGPPIGPPRGRTGDAGVGP	PPGPPGPPPSAGDFSF	LPQPPQEKADGGRY	1215
Qy	63	YRADDANVRRDRDLEVD	TTLKSLSQIENIRSP	EGSRKNPARTCDL	KMCHSDWKSGEYW 122
Db	1216	YRADDANVRRDRDLEVD	TTLKSLSQIENIRSP	EGSRKNPARTCDL	KMCHSDWKSGEYW 1275
Qy	123	IDNQCGLDAIKVFCNMET	GETCVPTQPSVAKQW	YISKPKDKRHW	FGESMTDGFQ 182
Db	1276	IDNQCGLDAIKVFCNMET	GETCVPTQPSVAKQW	YISKPKDKRHW	FGESMTDGFQ 1335
Qy	183	FYGGGSGSPADVAIQ	LTFLRLMSTRASQNI	TVHCKNSVAYMDQ	TGKLLKLLKGSNE 242
Db	1336	FYGGGSGSPADVAIQ	LTFLRLMSTRASQNI	TVHCKNSVAYMDQ	TGKLLKLLKGSNE 1395
Qy	243	IEIRAGNSRFTYSVTVD	GCTSHTGAWGKTIV	EYKTTKSSRLPI	IDVAPLDVGAPDQFPG 302
Db	1396	IEIRAGNSRFTYSVTVD	GCTSHTGAWGKTIV	EYKTTKSSRLPI	IDVAPLDVGAPDQFPG 1455
Qy	303	FDVGPVCF	L 311		


```
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-11-000-463-243

Query Match          99.1%; Score 1691; DB 7; Length 1464;
Best Local Similarity 99.4%; Pred. No. 1e-131;
Matches 307; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 NGLPGPIGPPGRGRTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPQKHAHDGGRY 62
Db 1156 NGLPGPIGPPGRGRTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPQKHAHDGGRY 1215

QY 63 YRADDANVVRDRDLEVDVTTLKSLSQOIENTIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 122
Db 1216 YRADDANVVRDRDLEVDVTTLKSLSQOIENTIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 1275

QY 123 IDPNOGNCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKHVWFGESMTDGFQ 182
Db 1276 IDPNOGNCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKHVWFGESMTDGFQ 1335

QY 183 FEYGGGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLQGSNE 242
Db 1336 FEYGGGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLQGSNE 1395

QY 243 IEIRAGNSRFTYSVTVDGCTSHGTGANGKTIVIEYKTKSSRLPIIDVAPLDVGPDOEFG 302
Db 1396 IEIRAGNSRFTYSVTVDGCTSHGTGANGKTIVIEYKTKSSRLPIIDVAPLDVGPDOEFG 1455

QY 303 FDVGPVCFL 311
Db 1456 FDVGPVCFL 1464

RESULT 5
US-10-821-234-1096
; Sequence 1096, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1096
; LENGTH: 1467
; TYPE: PRT
; ORGANISM: Homo sapiens
; ORGANISM: Homo sapiens
US-10-821-234-1096

Query Match          99.1%; Score 1691; DB 6; Length 1467;
Best Local Similarity 99.4%; Pred. No. 1e-131;
Matches 307; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 NGLPGPIGPPGRGRTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPQKHAHDGGRY 62
```

```
Db 1159 NGLPGPIGPPGRGRTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPQKHAHDGGRY 1218
QY 63 YRADDANVVRDRDLEVDVTTLKSLSQOIENTIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 122
Db 1219 YRADDANVVRDRDLEVDVTTLKSLSQOIENTIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 1278
QY 123 IDPNOGNCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKHVWFGESMTDGFQ 182
Db 1279 IDPNOGNCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKHVWFGESMTDGFQ 1338
QY 183 FEYGGGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLQGSNE 242
Db 1339 FEYGGGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLQGSNE 1398
QY 243 IEIRAGNSRFTYSVTVDGCTSHGTGANGKTIVIEYKTKSSRLPIIDVAPLDVGPDOEFG 302
Db 1399 IEIRAGNSRFTYSVTVDGCTSHGTGANGKTIVIEYKTKSSRLPIIDVAPLDVGPDOEFG 1458
QY 303 FDVGPVCFL 311
Db 1459 FDVGPVCFL 1467

RESULT 6
US-10-784-004-344
; Sequence 344, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 344
; LENGTH: 1453
; TYPE: PRT
; ORGANISM: rat
; ORGANISM: rat
US-10-784-004-344

Query Match          93.2%; Score 1591; DB 6; Length 1453;
Best Local Similarity 90.9%; Pred. No. 1.9e-123;
Matches 281; Conservative 16; Mismatches 12; Indels 0; Gaps 0;

QY 3 NGLPGPIGPPGRGRTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPQKHAHDGGRY 62
Db 1145 NGLPGPIGPPGRGRTGDAGVGPVGGPPGPPGPPSAGDFSFLLPQPQKHAHDGGRY 1204
QY 63 YRADDANVVRDRDLEVDVTTLKSLSQOIENTIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 122
Db 1205 YRADDANVVRDRDLEVDVTTLKSLSQOIENTIRSPGSRKNPARTCRDLKMKCHSDWKSGEYW 1264
QY 123 IDPNOGNCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKHVWFGESMTDGFQ 182
Db 1265 IDPNOGNCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKKHVWFGESMTDGFQ 1324
QY 183 FEYGGGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLQGSNE 242
Db 1325 FEYGGGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLQGSNE 1384
QY 243 IEIRAGNSRFTYSVTVDGCTSHGTGANGKTIVIEYKTKSSRLPIIDVAPLDVGPDOEFG 302
Db 1385 IEIRAGNSRFTYSVTVDGCTSHGTGANGKTIVIEYKTKSSRLPIIDVAPLDVGPDOEFG 1444
QY 303 FDVGPVCFL 311
Db 1445 MDIGPACFV 1453

RESULT 7
US-10-784-004-434
```

[illegible]

[illegible]

THIS PAGE LEFT BLANK

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	751	97.4	5921	6	CS833976	Q833976	Sequence
2	751	97.4	5921	6	CS063169	CS063169	Sequence
3	751	97.4	5931	6	CS080950	CS080950	Sequence
4	751	97.4	6728	6	AX167114	AX167114	Sequence
5	751	97.4	6728	6	AX330012	AX330012	Sequence
6	751	97.4	6728	6	AX330557	AX330557	Sequence
7	751	97.4	6728	6	AX333300	AX333300	Sequence
8	751	97.4	6728	6	AX393330	AX393330	Sequence
9	751	97.4	6728	6	AX411302	AX411302	Sequence
10	751	97.4	6728	8	HSPAL1CO	274615	H.sapiens m
11	746.2	96.8	4721	8	AB209597	AB209597	Homo sapi
12	746.2	96.8	6727	6	CO724221	CO724221	Sequence
13	745.8	96.7	3347	6	AX776246	AX776246	Sequence
14	745.8	96.7	3347	8	HUMCG1P1A1	K01228	Human proal
15	743	96.4	4752	8	BC036531	BC036531	Homo sapi
16	739	95.8	4778	6	AX806548	AX806548	Sequence
17	738.8	95.8	4409	6	A91174	A91174	Sequence 1
18	738.8	95.8	4409	6	BD070009	BD070009	Recombina

Db	3896	CCGTGACCTCAAGATGTCACCTCTGACTGGAAGTGGAGATCTGGATTGACCCCAA	3955	
Qy	197	CCAAGGTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGACCTG	256	
Db	3956	CCAAGGTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGACCTG	4015	
Qy	257	CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA	316	
Db	4016	CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA	4075	
Qy	317	GGACAAGAGGCGATGTCTGGTTCCGCGAGAGCATGCGATGGATTCAGTTCCAGTATGG	376	
Db	4076	GGACAAGAGGCGATGTCTGGTTCCGCGAGAGCATGCGATGGATTCAGTTCCAGTATGG	4135	
Qy	377	CGGCCAGGGCTCCGACCTGCCATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC	436	
Db	4136	CGGCCAGGGCTCCGACCTGCCATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC	4195	
Qy	437	CACGAGGCGCTCCAGAACATCACCTACCACTGCAAGAACAGCGTGGCTTACATGGACCA	496	
Db	4196	CACGAGGCGCTCCAGAACATCACCTACCACTGCAAGAACAGCGTGGCTTACATGGACCA	4255	
Qy	497	GCAGACTGGCAACCTCAAGAAGGCCCTGCTCCTCAAGGGCTCCAAAGAGATCGAGATCCG	556	
Db	4256	GCAGACTGGCAACCTCAAGAAGGCCCTGCTCCTCAAGGGCTCCAAAGAGATCGAGATCCG	4315	
Qy	557	CGCGAGGGCAACAGCGGCTTACCTACAGCGTCACTGTGCGTGGCTTACATGGACCA	616	
Db	4316	CGCGAGGGCAACAGCGGCTTACCTACAGCGTCACTGTGCGTGGCTTACATGGACCA	4375	
Qy	617	CGAGGCTGGGGCAAGACAGTGAATTGAATCAAAACCAACCAAGTCTCCCGCCTGCCAT	676	
Db	4376	CGAGGCTGGGGCAAGACAGTGAATTGAATCAAAACCAACCAAGTCTCCCGCCTGCCAT	4435	
Qy	677	CATCGATGTGGCCCTTGGACGTGTGGTGGCCAGACCCAGGAATTCGGCTTCGACGTTGG	736	
Db	4436	CATCGATGTGGCCCTTGGACGTGTGGTGGCCAGACCCAGGAATTCGGCTTCGACGTTGG	4495	
Qy	737	CCCTGTCTGCTTCTGTAAACTCCCTCCATC	767	
Db	4496	CCCTGTCTGCTTCTGTAAACTCCCTCCATC	4526	
RESULT 2				
LOCUS	CS063169	5921 bp	DNA	linear
DEFINITION	Sequence 153 from Patent EP1522594.			
ACCESSION	CS063169			
VERSION	CS063169.1	GI:62816627		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	1			
AUTHORS	Munnes, M. and Bojar, H.			
TITLE	Methods and kits for investigating cancer			
JOURNAL	Patent: EP 1522594-A 153 13-APR-2005;			
	Bayer HealthCare AG (DE)			
FEATURES	Location/Qualifiers			
source	1..5921			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
	/db_xref="taxon:9606"			
ORIGIN				
Query Match	97.48;	Score 751;	DB 6;	Length 5921;
Best Local Similarity	100.0%;	Pred. No. 1.3e-111;		
Matches 751;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	17	TGATGCCAATGTGGTTCTGTCGACCGTGGACCTCGAGGTGGACACCAACCCCTCAAGAGCCTGAG	76	

Db	3776	TGATGCCAATGTGGTTCTGTCGACCGTGAACCTTCAGAGTGGACACCACTTCAAGAGCCTGAG	3835	
Qy	77	CCAGCAGATCGAGAACATCCGGAGCCCGAGAGGAGCCGCAAGAACCCCGCCGACCTG	136	
Db	3836	CCAGCAGATCGAGAACATCCGGAGCCCGAGAGGAGCCGCAAGAACCCCGCCGACCTG	3895	
Qy	137	CGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTCTGGATTGACCCCAA	196	
Db	3896	CGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTCTGGATTGACCCCAA	3955	
Qy	197	CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGACCTG	256	
Db	3956	CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGACCTG	4015	
Qy	257	CGTGACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA	316	
Db	4016	CGTGACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA	4075	
Qy	317	GGACAAGAGGCGATGTCTGGTTCCGCGAGAGCATGCGATGGATTCAGTTCCAGTATGG	376	
Db	4076	GGACAAGAGGCGATGTCTGGTTCCGCGAGAGCATGCGATGGATTCAGTTCCAGTATGG	4135	
Qy	377	CGGCCAGGGCTCCGACCTGCCATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC	436	
Db	4136	CGGCCAGGGCTCCGACCTGCCATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC	4195	
Qy	437	CACGAGGCGCTCCAGAACATCACCTACCACTGCAAGAACAGCGTGGCTTACATGGACCA	496	
Db	4196	CACGAGGCGCTCCAGAACATCACCTACCACTGCAAGAACAGCGTGGCTTACATGGACCA	4255	
Qy	497	GCAGACTGGCAACCTCAAGAAGGCCCTGCTCCTCAAGGGCTCCAAAGAGATCGAGATCCG	556	
Db	4256	GCAGACTGGCAACCTCAAGAAGGCCCTGCTCCTCAAGGGCTCCAAAGAGATCGAGATCCG	4315	
Qy	557	CGCGAGGGCAACAGCGGCTTACCTACAGCGTCACTGTGCGTGGCTTACATGGACCA	616	
Db	4316	CGCGAGGGCAACAGCGGCTTACCTACAGCGTCACTGTGCGTGGCTTACATGGACCA	4375	
Qy	617	CGAGGCTGGGGCAAGACAGTGAATTGAATCAAAACCAACCAAGTCTCCCGCCTGCCAT	676	
Db	4376	CGAGGCTGGGGCAAGACAGTGAATTGAATCAAAACCAACCAAGTCTCCCGCCTGCCAT	4435	
Qy	677	CATCGATGTGGCCCTTGGACGTGTGGTGGCCAGACCCAGGAATTCGGCTTCGACGTTGG	736	
Db	4436	CATCGATGTGGCCCTTGGACGTGTGGTGGCCAGACCCAGGAATTCGGCTTCGACGTTGG	4495	
Qy	737	CCCTGTCTGCTTCTGTAAACTCCCTCCATC	767	
Db	4496	CCCTGTCTGCTTCTGTAAACTCCCTCCATC	4526	
RESULT 3				
LOCUS	CS080950	5921 bp	DNA	linear
DEFINITION	Sequence 153 from Patent WO2005040414.			
ACCESSION	CS080950			
VERSION	CS080950.1	GI:66348467		
KEYWORDS				
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	1			
AUTHORS	Munnes, M. and Bojar, H.			
TITLE	Methods and kits for investigating cancer			
JOURNAL	Patent: WO 2005040414-A 153 06-MAY-2005;			
	Bayer HealthCare AG (DE)			
FEATURES	Location/Qualifiers			
source	1..5921			
	/organism="Homo sapiens"			
	/mol_type="unassigned DNA"			
ORIGIN				
Query Match	97.48;	Score 751;	DB 6;	Length 5921;
Best Local Similarity	100.0%;	Pred. No. 1.3e-111;		
Matches 751;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	17	TGATGCCAATGTGGTTCTGTCGACCGTGGACCTCGAGGTGGACACCAACCCCTCAAGAGCCTGAG	76	

ORIGIN		/db_xref="taxon:9606"	
Query Match		97.4%; Score 751; DB 6; Length 5921;	
Best Local Similarity		100.0%; Pred. No. 1.3e-111;	
Matches 751; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	17	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTTCAGAGCCTGAG	76
Db	3776	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTTCAGAGCCTGAG	3835
Qy	77	CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG	136
Db	3836	CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG	3895
Qy	137	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA	196
Db	3896	CCGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTTGGATTGACCCCAA	3955
Qy	197	CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGAGACCTG	256
Db	3956	CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGAGACCTG	4015
Qy	257	CGTGTACCCCACTCAGCCCACTGTCGACCTCGAGGTGGAGAGTACTTGGATTGACCCCAA	316
Db	4016	CGTGTACCCCACTCAGCCCACTGTCGACCTCGAGGTGGAGAGTACTTGGATTGACCCCAA	4075
Qy	317	GGACAAAGAGGCTGCTGCTGCTCGGCGAGAGCATGACCGATGGATTCAGATTTCAGATATGG	376
Db	4076	GGACAAAGAGGCTGCTGCTGCTCGGCGAGAGCATGACCGATGGATTCAGATTTCAGATATGG	4135
Qy	377	CGGCGAGGCTCCGACCTCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATGTC	436
Db	4136	CGGCGAGGCTCCGACCTCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATGTC	4195
Qy	437	CACGAGGCTCCGAGAACATCAGCTACCTACCTGACCAAGAGAGGCTGACAGCA	496
Db	4196	CACGAGGCTCCGAGAACATCAGCTACCTGACCAAGAGAGGCTGACAGCA	4255
Qy	497	GCAGCTGGCAACTCAAGAGGCTGCTGCTCCTCAAGGCTCCAAAGAGATCGAGATCCG	556
Db	4256	GCAGCTGGCAACTCAAGAGGCTGCTGCTCCTCAAGGCTCCAAAGAGATCGAGATCCG	4315
Qy	557	CGCGAGGCAACAGCGCTTCACCTACAGCGTCACCTGTCGATGGCTGACAGTCAAC	616
Db	4316	CGCGAGGCAACAGCGCTTCACCTACAGCGTCACCTGTCGATGGCTGACAGTCAAC	4375
Qy	617	CGGAGCTGGGGCAAGACAGTGTGAATACAAACACCAAGTCTCCCGCTGCGCCAT	676
Db	4376	CGGAGCTGGGGCAAGACAGTGTGAATACAAACACCAAGTCTCCCGCTGCGCCAT	4435
Qy	677	CATCGATGTGCCCCCTTGGACGTTGTCGCCAGACAGGAAATTCGAGCTTCCAGCTTGG	736
Db	4436	CATCGATGTGCCCCCTTGGACGTTGTCGCCAGACAGGAAATTCGAGCTTCCAGCTTGG	4495
Qy	737	CCCTGTCTGCTTCTGTAATCTCCCTCCATC	767
Db	4496	CCCTGTCTGCTTCTGTAATCTCCCTCCATC	4526
RESULT 4		AX167114	
LOCUS		AX167114	
DEFINITION		Sequence 1 from Patent WO014455.	
ACCESSION		AX167114	
VERSION		AX167114.1	
KEYWORDS		GI:14596602	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
Hominidae; Homo.			
REFERENCE		1	

4136	Dd	CGGCCAGGGCTCCGACCGCTGCGGATGTGGCCATCCAGCTGACCTTCTTCGCGCCTGATGTC	4195
437	Qy	CACCGAGGCTCCCGAAGACATCACCTACCACTGCAAGAACAGCGTGGCCTTACATGAGACCA	496
4196	Dd	CACCGAGGCTCCCGAAGACATCACCTACCACTGCAAGAACAGCGTGGCCTTACATGAGACCA	4255
497	Qy	GCAGACTGGCAACCTCAAGAAGGCCCTTGCTCTCAAGGGCTCCAACGAGATCGAGATCCG	556
4256	Dd	GCAGACTGGCAACCTCAAGAAGGCCCTTGCTCTCAAGGGCTCCAACGAGATCGAGATCCG	4315
557	Qy	CGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATGGTGGCAAGAGTCCAC	616
4316	Dd	CGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATGGTGGCAAGAGTCCAC	4375
617	Qy	CGGAGCTGGGGCAAGACAGTGAATACAAACCAACCAAGTCTCCGCGCTGCCCAT	676
4376	Dd	CGGAGCTGGGGCAAGACAGTGAATACAAACCAACCAAGTCTCCGCGCTGCCCAT	4435
677	Qy	CATCGATGTGGCCCCCTTGAGCGTTGGTGGCCCGACAGCAAGGAATTCGGCTTCGACGTTGG	736
4436	Dd	CATCGATGTGGCCCCCTTGAGCGTTGGTGGCCCGACAGCAAGGAATTCGGCTTCGACGTTGG	4495
737	Qy	CCCTGTCTGCTTCTGTAACTCCCTCCATC	767
4496	Dd	CCCTGTCTGCTTCTGTAACTCCCTCCATC	4526
RESULT 5			
LOCUS	AX330012	6728 bp	DNA linear PAT 09-JAN-2002
DEFINITION	Sequence 521 from Patent WO0194629.		
ACCESSION	AX330012		
VERSION	AX330012.1	GI:18102990	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G., Horrigan, S., Soppet, D.R. and Weaver, Z.		
TITLE	Cancer gene determination and therapeutic screening using signature gene sets		
JOURNAL	Patent: WO 0194629-A 521 13-DEC-2001;		
FEATURES	Avalon Pharmaceuticals (US)		
source	1. .6728		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	97.4%;	Score 751;	DB 6; Length 6728;
Best Local Similarity	100.0%;	Pred. No. 1.2e-111;	
Matches 751;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	17	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTCAAGAGCCTGAG	76
Dd	3776	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTCAAGAGCCTGAG	3835
Qy	77	CCAGCAGATCGAAGACATCCGAGGCCAGAGGAGCGCAAGAACCCCGCCGACCTG	136
Dd	3836	CCAGCAGATCGAAGACATCCGAGGCCAGAGGAGCGCGAAGAACCCCGCCGACCTG	3895
Qy	137	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA	196
Dd	3896	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA	3955
Qy	197	CCAAGGCTGAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTAGACCTG	256
Dd	3956	CCAAGGCTGAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTAGACCTG	4015

257	Qy	CGTGTACCCCACTCAGCCCACTGTGGCCCAAGAGAACTGTGTACATCAGCAAGAACCCCA	316
4016	Dd	CGTGTACCCCACTCAGCCCACTGTGGCCCAAGAGAACTGTGTACATCAGCAAGAACCCCA	4075
317	Qy	GGACAAGAGGCATGTCTGGTTTCGGGGAGAGCATCAGCATGGATTCAGATTCGAGTATGG	376
4076	Dd	GGACAAGAGGCATGTCTGGTTTCGGGGAGAGCATCAGCATGGATTCAGATTCGAGTATGG	4135
377	Qy	CGGCCAGGGCTCCGAGCCCTGCGGATGTGGCCATCCAGCTGACCTTCTTGCGCCTGATGTC	436
4136	Dd	CGGCCAGGGCTCCGAGCCCTGCGGATGTGGCCATCCAGCTGACCTTCTTGCGCCTGATGTC	4195
437	Qy	CACCGAGGCTCCCAAGAACATCACTACCACTGCAAGAACAGCGTGGCCTTACATGAGACCA	496
4196	Dd	CACCGAGGCTCCCAAGAACATCACTACCACTGCAAGAACAGCGTGGCCTTACATGAGACCA	4255
497	Qy	GCAGACTGGCAACCTCAAGAAGGCCCTTGCTCTCAAGGGCTCCAACGAGATCGAGATCCG	556
4256	Dd	GCAGACTGGCAACCTCAAGAAGGCCCTTGCTCTCAAGGGCTCCAACGAGATCGAGATCCG	4315
557	Qy	CGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATGGTGGCAAGAGTCCAC	616
4316	Dd	CGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATGGTGGCAAGAGTCCAC	4375
617	Qy	CGGAGCTGGGGCAAGACAGTGAATACAAACCAACCAAGTCTCCGCGCTGCCCAT	676
4376	Dd	CGGAGCTGGGGCAAGACAGTGAATACAAACCAACCAAGTCTCCGCGCTGCCCAT	4435
677	Qy	CATCGATGTGGCCCCCTTGAGCGTTGGTGGCCCGACAGCAAGGAATTCGGCTTCGAGCTGG	736
4436	Dd	CATCGATGTGGCCCCCTTGAGCGTTGGTGGCCCGACAGCAAGGAATTCGGCTTCGAGCTGG	4495
737	Qy	CCCTGTCTGCTTCTGTAACTCCCTCCATC	767
4496	Dd	CCCTGTCTGCTTCTGTAACTCCCTCCATC	4526
RESULT 6			
LOCUS	AX330557	6728 bp	DNA linear PAT 09-JAN-2002
DEFINITION	Sequence 1066 from Patent WO0194629.		
ACCESSION	AX330557		
VERSION	AX330557.1	GI:18103535	
KEYWORDS	Homo sapiens (human)		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	1		
AUTHORS	Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G., Horrigan, S., Soppet, D.R. and Weaver, Z.		
TITLE	Cancer gene determination and therapeutic screening using signature gene sets		
JOURNAL	Patent: WO 0194629-A 1066 13-DEC-2001;		
FEATURES	Avalon Pharmaceuticals (US)		
source	1. .6728		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	97.4%;	Score 751;	DB 6; Length 6728;
Best Local Similarity	100.0%;	Pred. No. 1.2e-111;	
Matches 751;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;
Qy	17	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTCAAGAGCCTGAG	76
Dd	3776	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTCAAGAGCCTGAG	3835
Qy	77	CCAGCAGATCGAAGACATCCGAGGCCAGAGGAGCGCAAGAACCCCGCCGACCTG	136

Db 3836 CCAGCAGATCGAAGAACATCCGAGGCCCGAGAGGAGCGGCAAGAACCCCGCCGCCACCTG 3895
Qy 137 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3896 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955
Qy 197 CCAGAGGTCGAACCTGGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGCTG 256
Db 3956 CCAAGGCTGCAACCTGGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGCTG 4015
Qy 257 CGTGTACCCCACTCAGCCCACTGGCGGAGAGCATGACCCGATGATG 316
Db 4016 CGTGTACCCCACTCAGCCCACTGGCGGAGAGCATGACCCGATGATG 4075
Qy 317 GGACAAGAGGAGCATGCTGGTTCGGCGAGAGCATGACCCGATGATG 376
Db 4076 GGACAAGAGGAGCATGCTGGTTCGGCGAGAGCATGACCCGATGATG 4135
Qy 377 CGGCGAGGCTCCGAGCCCTGGGATGGCCATCCAGCTGACCTTCTGGCCCTGATGTC 436
Db 4136 CGGCGAGGCTCCGAGCCCTGGGATGGCCATCCAGCTGACCTTCTGGCCCTGATGTC 4195
Qy 437 CACCGAGGCTCCGAGAACATCACTTACCCTGCAAGAACAGAGTGGCTTACATGGACCA 496
Db 4196 CACCGAGGCTCCGAGAACATCACTTACCCTGCAAGAACAGAGTGGCTTACATGGACCA 4255
Qy 497 GCAGACTGGCAACTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556
Db 4256 GCAGACTGGCAACTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 4315
Qy 557 CGCGAGGGCAAGAGCGCTTACCTACAGGTCACCTGTCGATGGCTGCAAGAGTCAAC 616
Db 4316 CGCGAGGGCAAGAGCGCTTACCTACAGGTCACCTGTCGATGGCTGCAAGAGTCAAC 4375
Qy 617 CGAGACTGGCGGCAAGAGTGAATGAATACAAACCAAGTCCCTCCGCTGCCCCAT 676
Db 4376 CGAGACTGGCGGCAAGAGTGAATGAATACAAACCAAGTCCCTCCGCTGCCCCAT 4435
Qy 677 CATCGATGGGCGGCAAGAGTGAATGAATACAAACCAAGTCCCTCCGCTGCCCCAT 736
Db 4436 CATCGATGGGCGGCAAGAGTGAATGAATACAAACCAAGTCCCTCCGCTGCCCCAT 4495
Qy 737 CCCTGTCTGCTCTCTGTAACCTCCCTCCATC 767
Db 4496 CCCTGTCTGCTCTCTGTAACCTCCCTCCATC 4526

RESULT 7
AX333300
LOCUS AX333300
DEFINITION Sequence 3809 from Patent WO0194629.
ACCESSION AX333300
VERSION AX333300.1 GI:18123934
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1
REFERENCE
AUTHORS Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,
Horrigan, S., Soppet, D.R. and Weaver, Z.
TITLE Cancer gene determination and therapeutic screening using signature
gene sets
JOURNAL Patent: WO 0194629-A 3809 13-DEC-2001;
Avalon Pharmaceuticals (US)
FEATURES
source Location/Qualifiers
1 . 6728
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN

Query Match 97.4%; Score 751; DB 6; Length 6728;
Best Local Similarity 100.0%; Pred. No. 1.2e-111;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGTGTTCTGTCACCGTGTGACCTCGAGGTGGACACACACCTCTCAAGAGCTTGAG 76
Db 3776 TGATGCCAATGTGTGTTCTGTCACCGTGTGACCTCGAGGTGGACACACACCTCTCAAGAGCTTGAG 3835
Qy 77 CCAGCAGATCGAAGAACATCCGAGGCCCGAGAGGAGCCGCAAGAACCCCGCCGCCACCTG 136
Db 3836 CCAGCAGATCGAAGAACATCCGAGGCCCGAGAGGAGCCGCAAGAACCCCGCCGCCACCTG 3895
Qy 137 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3896 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955
Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGCTG 256
Db 3956 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGCTG 4015
Qy 257 CGTGTACCCCACTCAGCCCACTGGCGGAGAGCATGCTGTCATCAAGAACAGAACCCCAA 316
Db 4016 CGTGTACCCCACTCAGCCCACTGGCGGAGAGCATGCTGTCATCAAGAACAGAACCCCAA 4075
Qy 317 GGACAAGAGGAGCATGCTGGTTCGGCGAGAGCATGACCCGATGATG 376
Db 4076 GGACAAGAGGAGCATGCTGGTTCGGCGAGAGCATGACCCGATGATG 4135
Qy 377 CGGCGAGGCTCCGAGCCCTGGGATGGCCATCCAGCTGACCTTCTGGCCCTGATGTC 436
Db 4136 CGGCGAGGCTCCGAGCCCTGGGATGGCCATCCAGCTGACCTTCTGGCCCTGATGTC 4195
Qy 437 CACCGAGGCTCCGAGAACATCACTTACCCTGCAAGAACAGAGTGGCTTACATGGACCA 496
Db 4196 CACCGAGGCTCCGAGAACATCACTTACCCTGCAAGAACAGAGTGGCTTACATGGACCA 4255
Qy 497 GCAGACTGGCAACTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556
Db 4256 GCAGACTGGCAACTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 4315
Qy 557 CGCGAGGGCAAGAGCGCTTACCTACAGGTCACCTGTCGATGGCTGCAAGAGTCAAC 616
Db 4316 CGCGAGGGCAAGAGCGCTTACCTACAGGTCACCTGTCGATGGCTGCAAGAGTCAAC 4375
Qy 617 CGAGACTGGGCGGCAAGAGTGAATGAATACAAACCAAGTCCCTCCGCTGCCCCAT 676
Db 4376 CGAGACTGGGCGGCAAGAGTGAATGAATACAAACCAAGTCCCTCCGCTGCCCCAT 4435
Qy 677 CATCGATGGGCGGCAAGAGTGAATGAATACAAACCAAGTCCCTCCGCTGCCCCAT 736
Db 4436 CATCGATGGGCGGCAAGAGTGAATGAATACAAACCAAGTCCCTCCGCTGCCCCAT 4495
Qy 737 CCCTGTCTGCTCTCTGTAACCTCCCTCCATC 767
Db 4496 CCCTGTCTGCTCTCTGTAACCTCCCTCCATC 4526

RESULT 8
AX393330
LOCUS AX393330
DEFINITION Sequence 260 from Patent WO0210217.
ACCESSION AX393330
VERSION AX393330.1 GI:19701337
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1
REFERENCE
AUTHORS St Croix, B., Kinzler, K.W. and Vogelstein, B.
TITLE Endothelial cell expression patterns
JOURNAL Patent: WO 0210217-A 260 07-FEB-2002;

FEATURES		The Johns Hopkins University (US)		Homo sapiens	
source		Location/Qualifiers		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.	
ORIGIN		1..6728		1	
		/organism="Homo sapiens"		Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.	
		/mol_type="unassigned DNA"		Gene expression profiles in liver cancer	
		/db_xref="taxon:9606"		Patent: WO 0229103-A 3949 11-APR-2002;	
				GENE LOGIC INC (US)	
Query Match		97.4%; Score 751; DB 6; Length 6728;			
Best Local Similarity		100.0%; Pred. No. 1.2e-111;			
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	17	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTTCAAGAGCCTGAG	76	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTTCAAGAGCCTGAG	76
Db	3776	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTTCAAGAGCCTGAG	3835	TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCACTTCAAGAGCCTGAG	3835
QY	77	CCAGCAGATCGAGAACATCCGGAGCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG	136	CCAGCAGATCGAGAACATCCGGAGCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG	136
Db	3836	CCAGCAGATCGAGAACATCCGGAGCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG	3895	CCAGCAGATCGAGAACATCCGGAGCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG	3895
QY	137	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA	196	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA	196
Db	3896	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA	3955	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA	3955
QY	197	CCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG	256	CCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG	256
Db	3956	CCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG	4015	CCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG	4015
QY	257	CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA	316	CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA	316
Db	4016	CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA	4075	CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA	4075
QY	317	GGACAAGAGGCATGTCTGGTTTCGGCAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	376	GGACAAGAGGCATGTCTGGTTTCGGCAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	376
Db	4076	GGACAAGAGGCATGTCTGGTTTCGGCAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	4135	GGACAAGAGGCATGTCTGGTTTCGGCAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	4135
QY	377	CGGCCAGGGCTCCGACCTCGCATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC	436	CGGCCAGGGCTCCGACCTCGCATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC	436
Db	4136	CGGCCAGGGCTCCGACCTCGCATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC	4195	CGGCCAGGGCTCCGACCTCGCATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC	4195
QY	437	CACGAGGCTCCAGAACATCACTTACCTGCAAGAAACAGCGTGGCTTACATGGAACCA	496	CACGAGGCTCCAGAACATCACTTACCTGCAAGAAACAGCGTGGCTTACATGGAACCA	496
Db	4196	CACGAGGCTCCAGAACATCACTTACCTGCAAGAAACAGCGTGGCTTACATGGAACCA	4255	CACGAGGCTCCAGAACATCACTTACCTGCAAGAAACAGCGTGGCTTACATGGAACCA	4255
QY	497	GCAGACTGGGCAACCTCAAGAGCCCTGCTCCTCAAGGGCTCCAAGAGATCGAGATCCG	556	GCAGACTGGGCAACCTCAAGAGCCCTGCTCCTCAAGGGCTCCAAGAGATCGAGATCCG	556
Db	4256	GCAGACTGGGCAACCTCAAGAGCCCTGCTCCTCAAGGGCTCCAAGAGATCGAGATCCG	4315	GCAGACTGGGCAACCTCAAGAGCCCTGCTCCTCAAGGGCTCCAAGAGATCGAGATCCG	4315
QY	557	CGCCGAGGGCAACAGCCGCTTTCACCTACAGCGTCACTGTTCGATGGCTGCAACGATCACC	616	CGCCGAGGGCAACAGCCGCTTTCACCTACAGCGTCACTGTTCGATGGCTGCAACGATCACC	616
Db	4316	CGCCGAGGGCAACAGCCGCTTTCACCTACAGCGTCACTGTTCGATGGCTGCAACGATCACC	4375	CGCCGAGGGCAACAGCCGCTTTCACCTACAGCGTCACTGTTCGATGGCTGCAACGATCACC	4375
QY	617	CGGAGCCTGGGGCAAGACAGTGTGAATACAAAACCAACGATGCTCCCGCCTGCCCAT	676	CGGAGCCTGGGGCAAGACAGTGTGAATACAAAACCAACGATGCTCCCGCCTGCCCAT	676
Db	4376	CGGAGCCTGGGGCAAGACAGTGTGAATACAAAACCAACGATGCTCCCGCCTGCCCAT	4435	CGGAGCCTGGGGCAAGACAGTGTGAATACAAAACCAACGATGCTCCCGCCTGCCCAT	4435
QY	677	CATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCGAGAAATTCGGCTTCGACGTTGG	736	CATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCGAGAAATTCGGCTTCGACGTTGG	736
Db	4436	CATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCGAGAAATTCGGCTTCGACGTTGG	4495	CATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCGAGAAATTCGGCTTCGACGTTGG	4495
QY	737	CCCTGTCTGCTTCTGTGAAACTCCCTCCATC 767		CCCTGTCTGCTTCTGTGAAACTCCCTCCATC 767	
Db	4496	CCCTGTCTGCTTCTGTGAAACTCCCTCCATC 4526		CCCTGTCTGCTTCTGTGAAACTCCCTCCATC 4526	
RESULT 9					
AX411302					
LOCUS					
DEFINITION Sequence 3949 from Patent WO0229103.					
ACCESSION AX411302					
VERSION AX411302.1 GI:21444007					
KEYWORDS Homo sapiens (human)					
SOURCE					

exon	/number=22	197	CCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGGTGAGACCTG	256
exon	1635. .1733	3956	CCAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGGTGAGACCTG	4015
exon	1734. .1787	257	CGTGTAACCCACTCAGCCAGTGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA	316
exon	/number=24	4016	CGTGTAACCCACTCAGCCAGTGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA	4075
exon	1788. .1886	317	GGACAAGAGGCATGTCTGGTTCGGCGAGAGCATGCCGATGGATTCACAGTTCGAGTATGG	376
exon	/number=25	4076	GGACAAGAGGCATGTCTGGTTCGGCGAGAGCATGCCGATGGATTCACAGTTCGAGTATGG	4135
exon	1887. .1940	377	CGGCCAGGGCTCCGACCCCTGCGGATGTGGCCATCCAGCTGACCTTCTCGGCTGATGC	4195
exon	/number=26	437	CAACGAGGCTCCAGAACATCACTTACAGCGCTCACTGTCTGATGGCTGACGAGTCCAC	496
exon	1941. .1994	4196	CAACGAGGCTCCAGAACATCACTTACAGCGCTCACTGTCTGATGGCTGACGAGTCCAC	4255
exon	/number=27	497	GCAGACTGGCAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGTCCGATCCG	556
exon	1995. .2048	4256	GCAGACTGGCAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGTCCGATCCG	4315
exon	/number=28	557	CGCCGAGGCAACAGCGCTTCACTTACAGCGCTCACTGTCTGATGGCTGACGAGTCCAC	616
exon	2049. .2102	4316	CGCCGAGGCAACAGCGCTTCACTTACAGCGCTCACTGTCTGATGGCTGACGAGTCCAC	4375
exon	/number=29	617	CGGAGCTGGGGCAAGACAGTGAATTAACAAACCAACAGTCTCCGCTGCGCCAT	676
exon	2103. .2147	4376	CGGAGCTGGGGCAAGACAGTGAATTAACAAACCAACAGTCTCCGCTGCGCCAT	4435
exon	/number=30	677	CATCGATGTGGCCCTTGGAGCTTGGTGGCCAGACAGGAATTCGGCTTCGACCTTGG	736
exon	2148. .2246	4436	CATCGATGTGGCCCTTGGAGCTTGGTGGCCAGACAGGAATTCGGCTTCGACCTTGG	4495
exon	/number=31	737	CCCTGTCTCTCTCTGTAACCTCCCTCCATC	767
exon	2247. .2354	4496	CCCTGTCTCTCTCTGTAACCTCCCTCCATC	4526
exon	/number=32			
exon	2355. .2462			
exon	/notes="exons 33-34"			
exon	2463. .2516			
exon	/number=35			
exon	2517. .2570			
exon	/number=36			
exon	2571. .2678			
exon	/number=37			
exon	2679. .2732			
exon	/number=38			
exon	2733. .2786			
exon	/number=39			
exon	2787. .2948			
exon	/number=40			
exon	2949. .3056			
exon	/number=41			
exon	3057. .3164			
exon	/number=42			
exon	3165. .3218			
exon	/number=43			
exon	3219. .3326			
exon	/number=44			
exon	3327. .3380			
exon	/number=45			
exon	3381. .3488			
exon	/number=46			
exon	3489. .3542			
exon	/number=47			
exon	3543. .3650			
exon	/number=48			
old_sequence	3569. .3571			
	/citation=[3]			
	/replace="cnc"			
exon	3651. .3933			
misc_feature	/number=49			
	3696. .3773			
Query Match 97.4%; Score 751; DB 8; Length 6728;				
Best Local Similarity 100.0%; Pred. NO. 1.2e-111;				
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	17	TGATGCCAATGTGGTTCGTTGACCGTGACCTCGAGGTGGACACCACTTCAAGAGCTTGAG	76	
Db	3776	TGATGCCAATGTGGTTCGTTGACCGTGACCTCGAGGTGGACACCACTTCAAGAGCTTGAG	3835	
Qy	77	CCAGCAGATCGAAGACATCCGGAGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG	136	
Db	3836	CCAGCAGATCGAAGACATCCGGAGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG	3895	
Qy	137	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA	196	
Db	3896	CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA	3955	

RESULT 11
AB209597
LOCUS Homo sapiens mRNA for Collagen alpha 1 chain precursor variant
DEFINITION Homo sapiens mRNA for Collagen alpha 1 chain precursor variant
ACCESSION AB209597
VERSION AB209597.1 GI:62088773
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Totoki.Y., Toyoda.A., Takeda.T., Sakaki.Y., Tanaka.A., Yokoyama.S.,
Ohara.O., Nagase.T. and F.Kikuno.R.
TITLE None
JOURNAL Published Only in Database (2005)
REFERENCE 2 (bases 1 to 4721)
AUTHORS Totoki.Y., Toyoda.A., Takeda.T., Sakaki.Y., Tanaka.A., Yokoyama.S.,
Ohara.O., Nagase.T. and F.Kikuno.R.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2005) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 2-6-7 Kazusa-kamatari, Kisarazu,
Chiba, 292-0818, Japan [E-mail:cdnainfo@kazusa.or.jp,
URL:http://protein.gsc.riken.go.jp/, Tel:81-438-52-3930,
Fax:81-438-52-3931]
COMMENT This work was supported in part by the National Project on Protein
Structural and Functional Analysis, Ministry of Education, Culture
, Sports, Science and Technology of Japan. Totoki Y, Toyoda A,
Takeda T, Sakaki Y, Tanaka A, Yokoyama S. RIKEN Genomic Sciences

Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.

e-mail: aktanaka@postman.riken.go.jp

URL: http://protein.gsc.riken.go.jp/.

FEATURES

Location/Qualifiers
1..4721
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="sj06448"
/issue_type="spleen"
/note="this clone is also named as hsk003001786
vector:pBluescriptII SK plus"
<1..4721
/gene="Collagen alpha 1 chain precursor variant"
<93..4496
/gene="Collagen alpha 1 chain precursor variant"
/note="Start codon is not identified."
/codon_start=1
/evidence=not experimental
/product="Collagen alpha 1 chain precursor variant"
/protein_id="BAD92834.1"
/db_xref="GI:62088774"
/translation="GLDMFSFVDLRLLALLTHGQEEQVEQDDEIPITCV
QNGLRVHDRVWPKPCRIICVDNKGKLVCDIVICDTEKNCPGAEPVEGECPCVPDGS
ESPTDQETTGVEGKGTGIRGPRGPPAGPRGDIQPGCLPGPPGPPGPPGLGG
NFAPQLSYDESTGSIIVPGPMGSPRGLPGPCAPQFQFPPGPRGPRGASG
FMGPRGPPGKNGDDGEAKGPKRGERPPGPGARGLPGTAGLPGMKHGRGFSGL
DGAKGDAGPAGKPGSPGNGAPGQMGPRGLPGERGPGAPGAPGAGNDATGAA
GPGPTGAPGPPGPAVAGKAGAPQPGRSGPGQVRGEPGPPGAGAGPAGNPG
ADGPGAKGANGAPAGAPGPPGARGSPGQPGPPGPKNGSRGPPGPGSKGDTGA
KBPBGVVGVPQPGPAGKRGARGPPTGLPPGERGPGSGRFPAGDVGAPK
GPAGERSPPAGPKSGRAGRPGRAGLFGAKGLTSGPSGPPGKTIPTGPPAGQDG
RPGPPGARGAGVWGVPFGPKAAGEPGKAGERGVPPGPAVGAGKDGGAAGQP
PGPAGPAGERGEGSPGPGQGLPGPPAGPPGAGKPGEGQVPGDILGAPGSGRGER
LPGPRGDRGAPKGDGSPKQGVRLGTGPIGPPGAPAGPKDGSRSPGAPPTGA
RGAPDRGEPGPPGAPGPPGADGQPGAKGPRDAGKADGAPGPPGAPGPPPI
GNVAPGAKGARGSPGPPGATGFPGAAGVPPGPPGPNAGPPGPPGAGKGGPRG
ETGPAGRPGVPPGPPGAGKSGPGADGAPGAPGPPGAGRGGVGLPGRGE
RGFPGLPGSGEPGKPGSGASGERGPPGPMGPPGLAGPPGSGRGAEGSPGRD
GSPKAGDRGETGAPGAPGAPGAPGPPGPPGAGKSGDRGETGAPGAPGPPGARG
PAGPGPRGDKNGLPGPIGPPGPRGRTGAPGPPGPPGPPGPPGPPGPPGPPGPPG
PQPPQKADGGRTYRDDANVRDRLVDITLKSQLEINIRSPGSRKNPRT
RDKMHWFSMTDGFQFYGGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVA
PKDRHWYFESMTDGFQFYGGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVA
YMDQOTGLKALLQGSNEIIRAEGNSRFTYVTVDGCTSHTGAWGKTVIEYKTKK
TSRLPIIDVAPLDVGADQEFDFGVPVCFI"

ORIGIN

Query Match 96.8%; Score 746.2; DB 8; Length 4721;
Best Local Similarity 99.6%; Pred. No. 7.8e-111;
Matches 748; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGTTCTGTCACCGTGCCTCGAGTGGACACCCCTCAAGAGCCTGAG 76
Db 3758 TGATGCCAATGTGTTCTGTCACCGTGCCTCGAGTGGACACCCCTCAAGAGCCTGAG 3817
Qy 77 CCAGCAGATCGAGAACATCCGAGCCCGAGAGGAGCCGCAAGAACCCCGCCGACCTG 136
Db 3818 CCAGCAGATCGAGAACATCCGAGCCCGAGAGGAGCCGCAAGAACCCCGCCGACCTG 3877
Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3878 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 3937
Qy 197 CCAAGGCTGCAACTGGATGTCATCAAAAGTCTTCTGCAACATGAGAGCTGGTAGACCTG 256
Db 3938 CCAAGGCTGCAACTGGATGTCATCAAAAGTCTTCTGCAACATGAGAGCTGGTAGACCTG 3997
Qy 257 CGTGTACCCCACTCAGCCAGTGTGGCCGAGAGAACTGGTACTATCAGCAGAACCCCAA 316
Db 3998 CGTGTACCCCACTCAGCCAGTGTGGCCGAGAGAACTGGTACTATCAGCAGAACCCCAA 4057

Qy 317 GGACAGAGGCATGTCTGGTTTCGGCGAGAGCATGACCGATGGATTCAGTTCGAGTATGG 376
Db 4058 GGACAGAGGCATGTCTGGTTTCGGCGAGAGCATGACCGATGGATTCAGTTCGAGTATGG 4117
Qy 377 CGGCCAGGGCTCCGACCCCTCGATGTGGCCATTCACAGCTGACCTTCCTGCGCCTGATGTC 436
Db 4118 CGGCCAGGGCTCCGACCCCTCGATGTGGCCATTCACAGCTGACCTTCCTGCGCCTGATGTC 4177
Qy 437 CACGAGGCTCCCGAGAACATCACCTACCTGCAAGAAACAGCGTGGCCCTACATGGACCA 496
Db 4178 CACGAGGCTCCCGAGAACATCACCTACCTGCAAGAAACAGCGTGGCCCTACATGGACCA 4237
Qy 497 GCAGACTGGCAACCTCAAGAGGCCCTGCTCTCCAGAGGCTCCAAAGAGATCGAGATCCG 556
Db 4238 GCAGACTGGCAACCTCAAGAGGCCCTGCTCTCCAGAGGCTCCAAAGAGATCGAGATCCG 4297
Qy 557 CGCCGAGGGCAACAGCGGCTTACCTACAGCGTCACTGTGATGGGTGACAGAGTCAAC 616
Db 4298 CGCCGAGGGCAACAGCGGCTTACCTACAGCGTCACTGTGATGGGTGACAGAGTCAAC 4357
Qy 617 CGAGGCTGGGGCAAGACAGTATGAATACAAACACCAAGTCTCTCCGCTGCCCAT 676
Db 4358 CGAGGCTGGGGCAAGACAGTATGAATACAAACACCAAGTCTCTCCGCTGCCCAT 4417
Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGGCCCGAGACCGAGATTCGGCTTCGACGTTGG 736
Db 4418 CATCGATGTGGCCCTTGGACGTTGGTGGCCCGAGACCGAGATTCGGCTTCGACGTTGG 4477
Qy 737 CCCTGTCTGCTTCTCTGTAACCTCCCTCCATC 767
Db 4478 CCCTGTCTGCTTCTCTGTAACCTCCCTCCATC 4508
RESULT 12
LOCUS CQ724221 5727 bp DNA linear PAT 03-FEB-2004
DEFINITION Sequence 10155 from Patent WO02068579.
ACCESSION CQ724221
VERSION CQ724221.1 GI:42285078
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1
AUTHORS Venter, C.J., Adams, M.C., Li, P.W. and Myers, E.W.
TITLE Kits, such as nucleic acid arrays, comprising a majority of
humanexons or transcripts, for detecting expression and other uses
thereof
JOURNAL Patent: WO 02068579-A 10155 06-SEP-2002;
PE Corporation (NY) (US)
FEATURES
source Location/Qualifiers
1..6727
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 96.8%; Score 746.2; DB 6; Length 6727;
Best Local Similarity 99.6%; Pred. No. 7.3e-111;
Matches 748; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGTTCTGTCACCGTGCCTCGAGTGGACACCCCTCAAGAGCCTGAG 76
Db 3776 TGATGCCAATGTGTTCTGTCACCGTGCCTCGAGTGGACACCCCTCAAGAGCCTGAG 3835
Qy 77 CCAGCAGATCGAGAACATCCGAGCCCGAGAGGAGCCGCAAGAACCCCGCCGACCTG 136
Db 3836 CCAGCAGATCGAGAACATCCGAGCCCGAGAGGAGCCGCAAGAACCCCGCCGACCTG 3895
Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAAGATGGAGAGTACTGGATTGACCCCAA 196

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.
OM nucleic - nucleic search, using sw model
Run on: April 27, 2006, 17:14:00 ; Search time 481.986 Seconds
(without alignments)
10661.049 Million cell updates/sec

Title: US-10-677-877A-3
Perfect score: 771
Sequence: 1 aagctacgtaagatctgat.....gtaaacctccctccatctaga 771

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21.*

- 1: Geneseqn1980s.*
- 2: Geneseqn1990s.*
- 3: Geneseqn2000s.*
- 4: Geneseqn2001as.*
- 5: Geneseqn2001bs.*
- 6: Geneseqn2002as.*
- 7: Geneseqn2002bs.*
- 8: Geneseqn2003as.*
- 9: Geneseqn2003bs.*
- 10: Geneseqn2003cs.*
- 11: Geneseqn2003ds.*
- 12: Geneseqn2004as.*
- 13: Geneseqn2004bs.*
- 14: Geneseqn2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	771	100.0	771	14	AEA28570 Human col
2	760	98.6	1947	14	Aea28582 Human sol
3	760	98.6	2294	14	Aea28574 Human alk
4	759	98.4	1542	14	Aea28578 Human sol
5	755	97.9	963	14	Aea28568 Human col
6	755	97.9	1734	14	Aea28576 Human sol
7	755	97.9	2139	14	Aea28580 Human sol
8	755	97.9	2487	14	Aea28572 Human alk
9	751	97.4	4213	13	AD97949 Protein f
10	751	97.4	5921	8	ACC50109 Breast ca
11	751	97.4	5921	11	Adp65048 Human col
12	751	97.4	5921	12	ADQ19469 Human sof
13	751	97.4	5921	13	Adg29586 Human col
14	751	97.4	5921	12	ADR99017 Collagen,
15	751	97.4	5921	14	Adz26572 Human typ
16	751	97.4	5921	14	Adz09709 Human bre
17	751	97.4	5921	14	Aea04387 Human cdn
18	751	97.4	6728	4	Aaf90491 Human pro
19	751	97.4	6728	6	ABL62184 Colon ade

20	751	97.4	6728	6	ABL62729	Abl62729 Colon ade
21	751	97.4	6728	6	ABL65472	Abl65472 Lung can
22	751	97.4	6728	6	ABL92119	Abl92119 Human Tum
23	751	97.4	6728	6	ABN97451	Abn97451 Gene #394
24	751	97.4	6728	6	ABV94755	Abv94755 Human pan
25	751	97.4	6728	8	ACC50108	Acc50108 Breast ca
26	751	97.4	6728	8	ACF12845	Acf12845 Human cer
27	751	97.4	6728	10	ADD14744	Add14744 Human src
28	751	97.4	6728	10	ABX72044	Abx72044 DNA encod
29	751	97.4	6728	11	ADP65585	Adp65585 Human mrn
30	751	97.4	6728	14	ADV70139	Adv70139 Tumox-ass
31	746.2	96.8	3232	10	AD887386	Ad887386 Human pan
32	746.2	96.8	4073	10	AD887389	Ad887389 Human pan
33	746.2	96.8	4744	10	ADJ56243	Adj56243 Human cdn
34	746.2	96.8	4770	4	AAS22441	Aas22441 Human cdn
35	746.2	96.8	4784	4	AAH98417	Aah98417 Murine ES
36	746.2	96.8	4791	4	AAS22677	Aas22677 Human cdn
37	746.2	96.8	4791	13	ADS98292	Ads98292 Protein f
38	746.2	96.8	4952	8	ACC46568	Acc46568 Human dit
39	746.2	96.8	4952	10	AD887382	Ad887382 Human pan
40	746.2	96.8	5017	10	AD887387	Ad887387 Human pan
41	746.2	96.8	5739	10	AD887383	Ad887383 Human pan
42	746.2	96.8	5919	12	ADF45431	Adf45431 Human vas
43	746.2	96.8	5920	8	ACA64815	Aca64815 Human col
44	746.2	96.8	5920	12	ADQ86616	Adq86616 Human tum
45	746.2	96.8	5920	13	ACN40246	Acn40246 Tumour-as

ALIGNMENTS

RESULT 1
AEA28570
ID AEA28570 standard; cdna; 771 BP.
XX
AC AEA28570;
XX

DT 28-JUL-2005 (first entry)

XX Human collagen alpha(I) C-propeptide T2 construct cdna.

DE protein secretion; fusion protein; protein activation; collagen I; ss;
KW gene.

XX Homo sapiens.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 12..755

FT /*tag= a

FT /partial

FT /product= "Human collagen alpha(I) C-propeptide T2

FT construct protein"

FT /note= "No start codon"

XX WO2005047850-A2.

XX 26-MAY-2005.

XX 04-OCT-2004; 2004WO-US032753.

XX 02-OCT-2003; 2003US-00677877.

XX (GENH-) GENHUNTER CORP.

XX Liang P;

XX WPI; 2005-386406/39.

XX P-PSDB; AEA28571.

XX Generating a secreted trimeric fusion protein by introducing into a cell

XX a DNA construct comprising a promoter linked to a template encoding a

XX signal peptide sequence followed by in-frame fusion to polypeptide to be

```
PT trimerized.
XX Claim 14; SEQ ID NO 3; 47pp; English.
XX
CC The invention relates to a novel method for generating a secreted
CC trimeric fusion protein. The method comprises creating a DNA construct
CC comprising a transcriptional promoter linked to a template encoding a
CC signal peptide sequence followed, via an in-frame fusion, to the
CC polypeptide to be trimerized, which in turn is joined in-frame to a
CC polypeptide capable of self-trimerization which is heterologous to the
CC first polypeptide to be trimerized, and introducing the construct into a
CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
CC only the extracellular ligand-binding domain, may be used to intercept a
CC ligand and thus overcome the overactivation of a receptor during disease
CC treatment. Trimeric receptor decoys theoretically should have a much
CC higher affinity to their ligand than their dimeric counterparts. Such
CC rationally designed soluble trimeric receptor analogs could significantly
CC increase clinical benefits, as well as lower the amount or frequency of
CC drug injections for each patient. The method of the invention may be
CC useful for generating a secreted trimeric fusion protein.
CC
CC more efficient neutralization of the biological activities of their
CC naturally occurring trimeric ligands. The current sequence is that of the
CC human collagen alpha(I) C-propeptide T2 construct cDNA of the invention.
CC The construct includes a mutated BMP-1 recognition site but does not
CC contain the glycine-repeat triple helical region.
XX
SQ Sequence 771 BP; 186 A; 247 C; 204 G; 134 T; 0 U; 0 Other;

Query Match      100.0%; Score 771; DB 14; Length 771;
Best Local Similarity 100.0%; Pred. No. 2.6e-162;
Matches 771; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  AAGCTTACGTAAGTCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACCA 60
DB 1  AAGCTTACGTAAGTCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACCA 60

QY 61  CCCTCAAGAGCCTGAGCCAGCAGATCGAACAATCCGGAGCCCGAGGGAAGCCGCAAGA 120
DB 61  CCCTCAAGAGCCTGAGCCAGCAGATCGAACAATCCGGAGCCCGAGGGAAGCCGCAAGA 120

QY 121 ACCCGGCCCGCACTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGGT 180
DB 121 ACCCGGCCCGCACTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGGT 180

QY 181 ACTGGATTGACCCCAACCAAGGCTGCAACTCGATGTCATCAAGTCTTCTGCAACATGG 240
DB 181 ACTGGATTGACCCCAACCAAGGCTGCAACTCGATGTCATCAAGTCTTCTGCAACATGG 240

QY 241 AGACTGGTGAAGACTGGCTGTACCCCACTCAGCCAGTGTGGCCCAAGAACTGGGTACA 300
DB 241 AGACTGGTGAAGACTGGCTGTACCCCACTCAGCCAGTGTGGCCCAAGAACTGGGTACA 300

QY 301 TCAGCAAGAACCCCAAGGACCAAGAGGCAATCTGTGGTTCGCGGAGAGATGACCGATGGAT 360
DB 301 TCAGCAAGAACCCCAAGGACCAAGAGGCAATCTGTGGTTCGCGGAGAGATGACCGATGGAT 360

QY 361 TCAGTTCGAGTATGGCGGCGCGGGCTCCGACCTCGCGATGTCGCGATCCAGTGCACCT 420
DB 361 TCAGTTCGAGTATGGCGGCGCGGGCTCCGACCTCGCGATGTCGCGATCCAGTGCACCT 420

QY 421 TCCTGCGCCTGATGTCCACCGAGGCTCCCGAGAACATCACTACCACTGCAAGAACAGCG 480
DB 421 TCCTGCGCCTGATGTCCACCGAGGCTCCCGAGAACATCACTACCACTGCAAGAACAGCG 480

QY 481 TGGCCTACATGGACCAAGCAGATGGCAACTCTCAAGAGGCGCTCTCTCAAGGGTCTCCA 540
DB 481 TGGCCTACATGGACCAAGCAGATGGCAACTCTCAAGAGGCGCTCTCTCAAGGGTCTCCA 540

QY 541 ACCGATCGAGATCCGCGCGGCGGCAACCGCGCTTCACCTACAGGTCACTGTGATG 600
DB 541 ACCGATCGAGATCCGCGCGGCGGCAACCGCGCTTCACCTACAGGTCACTGTGATG 600

QY 601 GCTGCAAGTCAACCGGAGCCTGGGGCAAGACAGTGTGAATACAAAACCAACCAAGT 660
```

```

601  GCTGCAAGTCAACCGGAGCCTGGGGCAAGACAGTGTGAATACAAAACCAACCAAGT 660
661  CTTCCCGCCTGCGCCATCATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCAAGT 720
661  CTTCCCGCCTGCGCCATCATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCAAGT 720
721  TGGCTTCGACGTTGGCCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 771
721  TGGCTTCGACGTTGGCCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 771

RESULT 2
AEA28582
ID AEA28582 standard; cDNA; 1947 BP.
XX
AC AEA28582;
XX
DT 28-JUL-2005 (first entry)
XX
DE Human soluble CD4-collagen alpha(I) C-propeptide T2 fusion cDNA.
XX
KW protein secretion; fusion protein; protein activation; collagen I; ss;
KW gene; CD4; gene fusion; mutant.
XX
OS Homo sapiens.
OS Synthetic.
OS Chimeric.

Key Location/Qualifiers
CDS 24..1931
FT /tag= a
FT /product= "Human soluble CD4-collagen alpha(I) C-
FT propeptide T2 fusion protein"
XX
PN WO2005047850-A2.
XX
PD 26-MAY-2005.
XX
PF 04-OCT-2004; 2004WO-US032753.
XX
PR 02-OCT-2003; 2003US-00677877.
XX
PA (GENH-) GENHUNTER CORP.
XX
PI Liang P;
XX
DR WPI; 2005-386406/39.
XX
DR P-PSDB; AEA28583.
XX
CC Generating a secreted trimeric fusion protein by introducing into a cell
CC a DNA construct comprising a promoter linked to a template encoding a
CC signal peptide sequence followed by in-frame fusion to polypeptide to be
CC trimerized.
XX
PS Claim 16; SEQ ID NO 15; 47pp; English.
XX
CC The invention relates to a novel method for generating a secreted
CC trimeric fusion protein. The method comprises creating a DNA construct
CC comprising a transcriptional promoter linked to a template encoding a
CC signal peptide sequence followed, via an in-frame fusion, to the
CC polypeptide to be trimerized, which in turn is joined in-frame to a
CC polypeptide capable of self-trimerization which is heterologous to the
CC first polypeptide to be trimerized, and introducing the construct into a
CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
CC only the extracellular ligand-binding domain, may be used to intercept a
CC ligand and thus overcome the overactivation of a receptor during disease
CC treatment. Trimeric receptor decoys theoretically should have a much
CC higher affinity to their ligand than their dimeric counterparts. Such
CC rationally designed soluble trimeric receptor analogs could significantly
CC increase clinical benefits, as well as lower the amount or frequency of
CC drug injections for each patient. The method of the invention may be
CC useful for generating a secreted trimeric fusion protein to be used for
```


1595 CTGAGCCAGCAGATCCAGAACATCCGAGGCCAGAGGGAGCGCAAGAACCCCGCCCGC 1654
132 ACCTGCGGTGACCTCAAGATGTCCACTCTGACTTGAAGAGTGGAGAGTACTGGATTGAC 191
1655 ACCTGCGGTGACCTCAAGATGTCCACTCTGACTTGAAGAGTGGAGAGTACTGGATTGAC 1714
192 CCCAACCAAGGCTGCCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAG 251
1715 CCCAACCAAGGCTGCCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAG 1774
252 ACCTGCGGTGACCTCAAGATGTCCACTCTGACTTGAAGAGTGGAGAGTACTGGATTGAC 311
1775 ACCTGCGGTGACCTCAAGATGTCCACTCTGACTTGAAGAGTGGAGAGTACTGGATTGAC 1834
312 CCCAACCAAGGCTGCCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAG 371
1835 CCCAACCAAGGCTGCCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAG 1894
372 TATGGCGGCCGAGGCTCCGAGCCCTGCGATGTGGCCATCCAGCTGACCTTCTGCGCCTG 431
1895 TATGGCGGCCGAGGCTCCGAGCCCTGCGATGTGGCCATCCAGCTGACCTTCTGCGCCTG 1954
432 ATGTCCACCGAGGCTCCGAGAACATCACTACCTGCAAGAACAGCGTGGCTTACATG 491
1955 ATGTCCACCGAGGCTCCGAGAACATCACTACCTGCAAGAACAGCGTGGCTTACATG 2014
492 GACACAGAGTGGACCTCAAGAGGCGCTCTCTCAAGGGCTCCAAAGAGATCGAG 551
2015 GACACAGAGTGGACCTCAAGAGGCGCTCTCTCAAGGGCTCCAAAGAGATCGAG 2074
552 ATCCGCGCCGAGGCTCCGAGAACATCACTACCTGCAAGAACAGCGTGGCTTACATG 611
2075 ATCCGCGCCGAGGCTCCGAGAACATCACTACCTGCAAGAACAGCGTGGCTTACATG 2134
612 CACACCGAGGCTCCGAGAACATCACTACCTGCAAGAACAGCGTGGCTTACATG 671
2135 CACACCGAGGCTCCGAGAACATCACTACCTGCAAGAACAGCGTGGCTTACATG 2194
672 CCATCATGATGTGGCGCCCTTGGAGCTGGTGGCCCGAGACAGGAATTCGGCTTCGAC 731
2195 CCATCATGATGTGGCGCCCTTGGAGCTGGTGGCCCGAGACAGGAATTCGGCTTCGAC 2254
732 GTTGGCCCTGTCTGCTTCTGTAACCTCCCTCCATCTAGA 771
2255 GTTGGCCCTGTCTGCTTCTGTAACCTCCCTCCATCTAGA 2294

RESULT 4

AEA28578
ID AEA28578 standard; cDNA; 1542 BP.
XX AEA28578;
AC AC
XX AC
DT 28-JUL-2005 (first entry)
XX Human soluble TNF-RII-collagen alpha(I) C-propeptide T2 fusion cDNA.
XX protein secretion; fusion protein; protein activation; collagen I; ss;
XX gene; Tumor necrosis factor; TNF-RII; gene fusion; mutant.
XX Homo sapiens.
OS Synthetic.
OS Chimeric.
XX Key Location/Qualifiers
FH 18..1526
FT /tag= a
FT /product= "Human soluble TNF-RII-collagen alpha(I) C-
FT propeptide T2 fusion protein"
XX WO2005047850-A2.
XX

PD 26-MAY-2005.
XX 04-OCT-2004; 2004WO-US032753.
XX 02-OCT-2003; 2003US-00677877.
XX (GENH-) GENHUNTER CORP.
XX Liang P;
XX WPI; 2005-386406/39.
XX P-PSDB; AEA28579.
XX
XX Generating a secreted trimeric fusion protein by introducing into a cell
XX a DNA construct comprising a promoter linked to a template encoding a
XX signal peptide sequence followed by in-frame fusion to polypeptide to be
XX trimerized.
XX
XX Claim 15; SEQ ID NO 11; 47pp; English.
XX
XX The invention relates to a novel method for generating a secreted
XX trimeric fusion protein. The method comprises creating a DNA construct
XX comprising a transcriptional promoter linked to a template encoding a
XX signal peptide sequence followed, via an in-frame fusion, to the
XX polypeptide to be trimerized, which in turn is joined in-frame to a
XX polypeptide capable of self-trimerization which is heterologous to the
XX first polypeptide to be trimerized, and introducing the construct into a
XX eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
XX only the extracellular ligand-binding domain, may be used to intercept a
XX ligand and thus overcome the overactivation of a receptor during disease
XX treatment. Trimeric receptor decoys theoretically should have a much
XX higher affinity to their ligand than their dimeric counterparts. Such
XX rationally designed soluble trimeric receptor analogs could significantly
XX increase clinical benefits, as well as lower the amount or frequency of
XX drug injections for each patient. The method of the invention may be
XX useful for generating a secreted trimeric fusion protein to be used for
XX more efficient neutralization of the biological activities of their
XX naturally occurring trimeric ligands. The current sequence is that of the
XX human soluble TNF-RII-human collagen alpha(I) C-propeptide T2 fusion DNA
XX (sTNF-RII-T2) of the invention. The collagen T2 construct includes a
XX mutated BMP-1 recognition site but does not contain the glycine-repeat
XX triple helical region.
XX
XX Sequence 1542 BP; 337 A; 522 C; 425 G; 258 T; 0 U; 0 Other;

Query Match 98.4%; Score 759; DB 14; Length 1542;
Best Local Similarity 100.0%; Pred. No. 1.4e-159; Mismatches 0; Indels 0; Gaps 0;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 13 GATCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCAACCTCAAGAGCC 72
DB 784 GATCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCAACCTCAAGAGCC 843
QY 73 TGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGAGGAAAGCCGCAAGAACCCCGCCGCA 132
DB 844 TGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGAGGAAAGCCGCAAGAACCCCGCCGCA 903
QY 133 CTTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACC 192
DB 904 CTTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACC 963
QY 193 CCACCAAGCGTGCACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGA 252
DB 964 CCACCAAGCGTGCACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGA 1023
QY 253 CTTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACC 312
DB 1024 CTTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACC 1083
QY 313 CCAGAGCAAGAGGCGATGTCTGGTTCGGCGAGAGCATGACCGATGATTCGATTCGAGT 372
DB 1084 CCAGAGCAAGAGGCGATGTCTGGTTCGGCGAGAGCATGACCGATGATTCGATTCGAGT 1143

QY 677 CATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCAAGAAATTCGGCTTCGACGTTGG 736
Db |||||||
QY 869 CATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCAAGAAATTCGGCTTCGACGTTGG 928
Db |||||||
QY 737 CCTGTCTGCTTCTGTAACCTCCCTCATCTAGA 771
Db |||||||
QY 929 CCTGTCTGCTTCTGTAACCTCCCTCATCTAGA 963
Db |||||||

RESULT 6
AEA28576
ID AEA28576 standard; cDNA; 1734 BP.
XX
AC AEA28576;
XX
DT 28-JUL-2005 (first entry)
XX
DE Human soluble TNF-RII-collagen alpha(I) C-propeptide T0 fusion cDNA.
XX
XX protein secretion; fusion protein; protein activation; collagen I; ss;
XX Gene; Tumor necrosis factor; TNF-RII; gene fusion.
XX
OS Homo sapiens.
OS Synthetic.
OS Chimeric.

XX Key Location/Qualifiers
FT CDS 18..1718
FT /*tag= a
FT /product= "Human soluble TNF-RII-collagen alpha(I) C-
FT propeptide T0 fusion protein"
XX
PN WO2005047850-A2.
XX
XX 26-MAY-2005.
XX 04-OCT-2004; 2004WO-US032753.
XX
XX 02-OCT-2003; 2003US-00677877.
XX
XX (GENH-) GENHUNTER CORP.
XX
XX Liang P;
XX
XX WPI; 2005-396406/39.
XX P-PSDB; AEA28577.
XX

Generating a secreted trimeric fusion protein by introducing into a cell a DNA construct comprising a promoter linked to a template encoding a signal peptide sequence followed by in-frame fusion to polypeptide to be trimerized.
Claim 15; SEQ ID NO 9; 47pp; English.

The invention relates to a novel method for generating a secreted trimeric fusion protein. The method comprises creating a DNA construct comprising a transcriptional promoter linked to a template encoding a signal peptide sequence followed, via an in-frame fusion, to the polypeptide to be trimerized, which in turn is joined in-frame to a polypeptide capable of self-trimerization which is heterologous to the first polypeptide to be trimerized, and introducing the construct into a eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of only the extracellular ligand-binding domain, may be used to intercept a ligand and thus overcome the overactivation of a receptor during disease treatment. Trimeric receptor decoys theoretically should have a much higher affinity to their ligand than their dimeric counterparts. Such rationally designed soluble trimeric receptor analogs could significantly increase clinical benefits, as well as lower the amount or frequency of drug injections for each patient. The method of the invention may be useful for generating a secreted trimeric fusion protein to be used for more efficient neutralization of the biological activities of their naturally occurring trimeric ligands. The current sequence is that of the human soluble TNF-RII-human collagen alpha(I) C-propeptide T0 fusion DNA

CC (stNF-RII-T0) of the invention. The collagen T0 construct includes a partial glycine-repeat triple helical region.
XX
SQ Sequence 1734 BP; 358 A; 598 C; 477 G; 301 T; 0 U; 0 Other;

Query Match 97.9%; Score 755; DB 14; Length 1734;
Best Local Similarity 100.0%; Pred. No. 1.1e-158;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TGAATGCAATGTGTTGTCGTCGACCTTCGAGTGGACACCAACCTCAAGAGCCTGAG 76
Db |||||||
QY 980 TGAATGCAATGTGTTGTCGTCGACCTTCGAGTGGACACCAACCTCAAGAGCCTGAG 1039
Db |||||||
QY 77 CCAGCAGATCGAGAACATCCGGAGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG 136
Db |||||||
QY 1040 CCAGCAGATCGAGAACATCCGGAGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG 1099
Db |||||||
QY 137 CCGTGACCTCAAGATGTCACCTCTGACTGGAGAGTGGAGAGTACTGATGACCCCAA 196
Db |||||||
QY 1100 CCGTGACCTCAAGATGTCACCTCTGACTGGAGAGTGGAGAGTACTGATGACCCCAA 1159
Db |||||||
QY 197 CCAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
Db |||||||
QY 1160 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 1219
Db |||||||
QY 257 CGTGATACCCCACTCAGCCCAAGTGTGGCCAGAGAACTGTGTATCATCAGCAAGAACCCCAA 316
Db |||||||
QY 1220 CGTGATACCCCACTCAGCCCAAGTGTGGCCAGAGAACTGTGTATCATCAGCAAGAACCCCAA 1279
Db |||||||
QY 317 GGACAAAGAGGCAATGTCTGGTTCCGCGAGAGCATACCGATGGATTCAGTTCGAGTATGG 376
Db |||||||
QY 1280 GGACAAAGAGGCAATGTCTGGTTCCGCGAGAGCATACCGATGGATTCAGTTCGAGTATGG 1339
Db |||||||
QY 377 CGGCCAGGGCTCCGACCCCTGCCGATGTGGCCATCCAGCTGACCTTCTGGGCTGATGTC 436
Db |||||||
QY 1340 CGGCCAGGGCTCCGACCCCTGCCGATGTGGCCATCCAGCTGACCTTCTGGGCTGATGTC 1399
Db |||||||
QY 437 CACCGAGGGCTCCCGAGAACATCACTTACCACTGCAAGAAACAGCGTGGCTTACATGGACCA 496
Db |||||||
QY 1400 CACCGAGGGCTCCCGAGAACATCACTTACCACTGCAAGAAACAGCGTGGCTTACATGGACCA 1459
Db |||||||
QY 497 GCAGACTGGCAACCTCAAGAGGCCCTGTCTCTCAAGGGCTCCACGAGATCGAGATCGG 556
Db |||||||
QY 1460 GCAGACTGGCAACCTCAAGAGGCCCTGTCTCTCAAGGGCTCCACGAGATCGAGATCGG 1519
Db |||||||
QY 557 CGCCGAGGGCAACAGCGGCTTACCTTACAGCGTCTGATGCGTGGCTGCACAGTCCAC 616
Db |||||||
QY 1520 CGCCGAGGGCAACAGCGGCTTACCTTACAGCGTCTGATGCGTGGCTGCACAGTCCAC 1579
Db |||||||
QY 617 CGGAGGCTGGGGCAAGACAGTGAATGAAATACAAAACACCAAGTCTCTCCGCTGCCAT 676
Db |||||||
QY 1580 CGGAGGCTGGGGCAAGACAGTGAATGAAATACAAAACACCAAGTCTCTCCGCTGCCAT 1639
Db |||||||
QY 677 CATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCAAGAAATTCGGCTTCGACGTTGG 736
Db |||||||
QY 1640 CATCGATGTGGCCCTTGGACGTTGGTGGCCAGACCAAGAAATTCGGCTTCGACGTTGG 1699
Db |||||||
QY 737 CCTGTCTGCTTCTGTAACCTCCCTCATCTAGA 771
Db |||||||
QY 1700 CCTGTCTGCTTCTGTAACCTCCCTCATCTAGA 1734
Db |||||||

RESULT 7
AEA28580
ID AEA28580 standard; cDNA; 2139 BP.
XX
XX AEA28580;
XX
XX 28-JUL-2005 (first entry)
XX
XX Human soluble CD4-collagen alpha(I) C-propeptide T0 fusion cDNA.
XX
XX protein secretion; fusion protein; protein activation; collagen I; ss;
XX

KW	gene; CD4; gene fusion.	
OS	Homo sapiens.	
OS	Synthetic.	
OS	Chimeric.	
XX	Key	Location/Qualifiers
PH	CDS	24..2123
FT		/*tag= a
FT		/product= "Human soluble CD4-collagen alpha(I) C-
FT		propeptide T0 fusion protein"
XX		
XX		
PN	W02005047850-A2.	
XX		
PD	26-MAY-2005.	
XX		
PF	04-OCT-2004; 2004WO-US032753.	
XX		
PR	02-OCT-2003; 2003US-00677877.	
XX		
PA	(GENH-) GENHUNTER CORP.	
XX		
PI	Liang P;	
XX		
XX	WPI; 2005-386406/39.	
DR	P-PSDB; AEA28572.	
XX		
PT	Generating a secreted trimeric fusion protein by introducing into a cell	
PT	a DNA construct comprising a promoter linked to a template encoding a	
PT	signal peptide sequence followed by in-frame fusion to polypeptide to be	
PT	trimerized.	
XX		
PS	Claim 16; SEQ ID NO 13; 47pp; English.	
XX		
CC	The invention relates to a novel method for generating a secreted	
CC	trimeric fusion protein. The method comprises creating a DNA construct	
CC	comprising a transcriptional promoter linked to a template encoding a	
CC	signal peptide sequence followed, via an in-frame fusion, to the	
CC	polypeptide to be trimerized, which in turn is joined in-frame to a	
CC	polypeptide capable of self-trimerization which is heterologous to the	
CC	first polypeptide to be trimerized, and introducing the construct into a	
CC	eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of	
CC	only the extracellular ligand-binding domain, may be used to intercept a	
CC	ligand and thus overcome the overactivation of a receptor during disease	
CC	treatment. Trimeric receptor decoys theoretically should have a much	
CC	higher affinity to their ligand than their dimeric counterparts. Such	
CC	rationally designed soluble trimeric receptor analogs could significantly	
CC	increase clinical benefits, as well as lower the amount or frequency of	
CC	drug injections for each patient. The method of the invention may be	
CC	useful for generating a secreted trimeric fusion protein to be used for	
CC	more efficient neutralization of the biological activities of their	
CC	naturally occurring trimeric ligands. The current sequence is that of the	
CC	human soluble CD4-human collagen alpha(I) C-propeptide T0 fusion DNA	
CC	(SCD4-T0) of the invention. The collagen T0 construct includes a partial	
XX	glycine-repeat triple helical region.	
XX		
SQ	Sequence 2139 BP; 518 A; 621 C; 590 G; 410 T; 0 U; 0 Other;	
	Query Match 97.9%; Score 755; DB 14; Length 2139;	
	Best Local Similarity 100.0%; Pred. No. 1.2e-158;	
	Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	17 TGATGCCAAATGTGGTTCGTGACCTGACCTCGAGGTGGACACACCCCTCAAGAGCCCTGAG 76	
DB		
DB	1385 TGATGCCAAATGTGGTTCGTGACCTGACCTCGAGGTGGACACACCCCTCAAGAGCCCTGAG 1444	
QY	77 CCAGCAGATCGAGAAACATCCGAGGCCCGAGAGGGAGCCGCAAGAACCCCGCCGACCTG 136	
DB		
DB	1445 CCAGCAGATCGAGAAACATCCGAGGCCCGAGAGGGAGCCGCAAGAACCCCGCCGACCTG 1504	
QY	137 CCGTGACCTCAGATGTGCCACTCTGACTGGAGAGTGGAGACTGAGTTGACCCCAA 196	
DB		
DB	1505 CCGTGACCTCAGATGTGCCACTCTGACTGGAGAGTGGAGACTGAGTTGACCCCAA 1564	
QY	197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 256	
DB		
DB	1565 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 1624	
QY	257 CGTGATACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316	
DB		
DB	1625 CGTGATACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 1684	
QY	317 GGACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCAGTTCCGATATGG 376	
DB		
DB	1685 GGACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCAGTTCCGATATGG 1744	
QY	377 CGGCGAGGGCTCCGACCCCTCGATGTGGCCATCCAGTGAACCTTCTGCGCCCTGATGTC 436	
DB		
DB	1745 CGGCGAGGGCTCCGACCCCTCGATGTGGCCATCCAGTGAACCTTCTGCGCCCTGATGTC 1804	
QY	437 CACCGAGGGCTCCGAGAAACATCACTACCACTCAAGAACAGCGTGGCCTTACATGGACCA 496	
DB		
DB	1805 CACCGAGGGCTCCGAGAAACATCACTACCACTCAAGAACAGCGTGGCCTTACATGGACCA 1864	
QY	497 GCAGACTGGCAACCTCAAGAGGCCCTGTCTCTCAAGGGCTCCAAACGAGATCGAGATCCG 556	
DB		
DB	1865 GCAGACTGGCAACCTCAAGAGGCCCTGTCTCTCAAGGGCTCCAAACGAGATCGAGATCCG 1924	
QY	557 CGCGGAGGGCAACAGCCGCTTACCTACAGCGTCACTGTGATGGCTGCGACGATCAAC 616	
DB		
DB	1925 CGCGGAGGGCAACAGCCGCTTACCTACAGCGTCACTGTGATGGCTGCGACGATCAAC 1984	
QY	617 CGGAGGCTGGGGCAAGACAGTGAATGAATACAAACACCAAGTCTCTCCGCGCTGCCAT 676	
DB		
DB	1985 CGGAGGCTGGGGCAAGACAGTGAATGAATACAAACACCAAGTCTCTCCGCGCTGCCAT 2044	
QY	677 CATCGATGTGGCCCTTGGACCTTGGTGGCCCGCAGACCAAGAAATTCGGCTTCGACGTTGG 736	
DB		
DB	2045 CATCGATGTGGCCCTTGGACCTTGGTGGCCCGCAGACCAAGAAATTCGGCTTCGACGTTGG 2104	
QY	737 CCTGTGCTGCTCTGCTGTAACCTCCCTCCATCTAGA 771	
DB		
DB	2105 CCTGTGCTGCTCTGCTGTAACCTCCCTCCATCTAGA 2139	
RESULT 8		
AEA28572		
ID	AEA28572 standard; cDNA; 2487 BP.	
XX		
AC	AEA28572;	
XX		
DT	28-JUL-2005 (first entry)	
XX		
DE	Human alkaline phosphatase-collagen alpha(I) C-propeptide fusion cDNA.	
XX		
KW	protein secretion; fusion protein; protein activation; collagen I; ss;	
KW	gene; Alkaline phosphatase; gene fusion.	
XX		
OS	Homo sapiens.	
OS	Synthetic.	
OS	Chimeric.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	12..2471
FT		/*tag= a
FT		/product= "Human alkaline phosphatase-collagen alpha(I) C
FT		-propeptide fusion protein"
XX		
PN	W02005047850-A2.	
XX		
PD	26-MAY-2005.	
XX		
PF	04-OCT-2004; 2004WO-US032753.	
XX		
XX	02-OCT-2003; 2003US-00677877.	
XX		

PA	(GENH-) GENHUNTER CORP.	
XX	Liang P;	
PI	WPI; 2005-386406/39.	
XX	P-PSDB; AEA28573.	
DR	Generating a secreted trimeric fusion protein by introducing into a cell	
XX	a DNA construct comprising a promoter linked to a template encoding a	
PT	signal peptide sequence followed by in-frame fusion to polypeptide to be	
PT	trimerized.	
XX	Claim 17; SEQ ID NO 5; 47pp; English.	
PS	The invention relates to a novel method for generating a secreted	
XX	trimeric fusion protein. The method comprises creating a DNA construct	
CC	comprising a transcriptional promoter linked to a template encoding a	
CC	signal peptide sequence followed, via an in-frame fusion, to the	
CC	polypeptide to be trimerized, which in turn is joined in-frame to a	
CC	polypeptide capable of self-trimerization which is heterologous to the	
CC	first polypeptide to be trimerized, and introducing the construct into a	
CC	eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of	
CC	only the extracellular ligand-binding domain, may be used to intercept a	
CC	ligand and thus overcome the overactivation of a receptor during disease	
CC	treatment. Trimeric receptor decoys theoretically should have a much	
CC	higher affinity to their ligand than their dimeric counterparts. Such	
CC	rationally designed soluble trimeric receptor analogs could significantly	
CC	increase clinical benefits, as well as lower the amount or frequency of	
CC	drug injections for each patient. The method of the invention may be	
CC	useful for generating a secreted trimeric fusion protein to be used for	
CC	more efficient neutralization of the biological activities of their	
CC	naturally occurring trimeric ligands. The current sequence is that of the	
CC	human placental alkaline phosphatase (AP)-human collagen alpha(I) C-	
CC	propeptide 10 fusion DNA (AP-10) of the invention. The collagen 10	
CC	construct includes a partial glycine-repeat triple helical region.	
XX	Sequence 2487 BP; 525 A; 813 C; 728 G; 421 T; 0 U; 0 Other;	
SQ	Query Match 97.9%; Score 755; DB 14; Length 2487;	
	Best Local Similarity 100.0%; Pred. No. 1.2e-158; Mismatches 0; Gaps 0;	
	Matches 755; Conservative 0; Indels 0; Gaps 0;	
QY	17 TGATGCCAATGTGGTGTGACCGGTGACCTCGAGGTGGACACCACTTCAAGAGCCTGAG 76	
DB	1733 TGATGCCAATGTGGTGTGACCGGTGACCTCGAGGTGGACACCACTTCAAGAGCCTGAG 1792	
QY	77 CCAGCAGATCGAGAACATCCGAGGCCCGAGGGAGCCGCAAGAACCCCGCCGACCTG 136	
DB	1793 CCAGCAGATCGAGAACATCCGAGGCCCGAGGGAGCCGCAAGAACCCCGCCGACCTG 1852	
QY	137 CCCTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 196	
DB	1853 CCCTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 1912	
QY	197 CCAAGGCTGCAACTCGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGCCTG 256	
DB	1913 CCAAGGCTGCAACTCGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGCCTG 1972	
QY	257 CGTGTACCCCACTCAGCCCACTGTGGCCGAGAGAACTGTATACAGCAAGAACCCCAA 316	
DB	1973 CGTGTACCCCACTCAGCCCACTGTGGCCGAGAGAACTGTGTATACAGCAAGAACCCCAA 2032	
QY	317 GGACAGAGCATGTCTGGTTCGGCGAGAGCATGACCGATGGATTCAGTTCGAGTATGG 376	
DB	2033 GGACAGAGCATGTCTGGTTCGGCGAGAGCATGACCGATGGATTCAGTTCGAGTATGG 2092	
QY	377 CGGCCAGGGTCCGACCTCGCATGTGGCCATCCAGCTGACCTTCTCGCGCCTGTATGTC 436	
DB	2093 CGGCCAGGGTCCGACCTCGCATGTGGCCATCCAGCTGACCTTCTCGCGCCTGTATGTC 2152	
QY	437 CACGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCGTGGCCCTACATGGACCA 496	
DB	2153 CACGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCGTGGCCCTACATGGACCA 2212	
QY	497 GCAGACTGGCAACCTCAAGAAGGCCCTGTCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556	
DB	2213 GCAGACTGGCAACCTCAAGAAGGCCCTGTCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 2272	
QY	557 CGCCGAGGGCAACAGCCGCTTACCTACAGCGTCACTGTCTGATGGCTGCACGAGTCAAC 616	
DB	2273 CGCCGAGGGCAACAGCCGCTTACCTACAGCGTCACTGTCTGATGGCTGCACGAGTCAAC 2332	
QY	617 CGGAGCCTCGGGCAACAGACAGTGAATGAATACAAAACCAAGTCTCTCCGCTCGCCAT 676	
DB	2333 CGGAGCCTCGGGCAACAGACAGTGAATGAATACAAAACCAAGTCTCTCCGCTCGCCAT 2392	
QY	677 CATCGATGTGGCCCTTGGACGTTGGTGGCCCGACAGACCAGGAATTCGGCTTCGACGTTGG 736	
DB	2393 CATCGATGTGGCCCTTGGACGTTGGTGGCCCGACAGACCAGGAATTCGGCTTCGACGTTGG 2452	
QY	737 CCTGTCTGTCTTCTGTAAACTCCCTCCATCTAGA 771	
DB	2453 CCTGTCTGTCTTCTGTAAACTCCCTCCATCTAGA 2487	
XX	RESULT 9	
AD	AD997949	
ID	AD997949 standard; DNA; 4213 BP.	
XX	AD997949;	
AC	30-DEC-2004 (first entry)	
DT	Protein factor discovery related isolated human DNA, SEQ ID No 213.	
DE	antiinflammatory; cytostatic; antimicrobial; gene therapy; inflammation;	
DE	leukemia; nervous system disorder; infection; gene; ds.	
KW	Homo sapiens.	
KW	WO2004087874-A2.	
XX	14-OCT-2004.	
XX	24-MAR-2004; 2004WO-US009202.	
PF	28-MAR-2003; 2003US-0458824P.	
XX	(NUVE-) NUVELO INC.	
XX	(DRMA/) DRMANAC R T.	
PA	Tang YT, Zhou P, Wang J, Wang ZW, Hu T;	
XX	WPI; 2004-737686/72.	
PI	P-PSDB; AD998184.	
DR	New polynucleotides encoding a polypeptide with biological activity, or	
PT	infectious.	
PT	Claim 1; SEQ ID NO 213; 253pp; English.	
XX	The invention relates to a novel isolated polynucleotide comprising any	
CC	of the 235 nucleotide sequences described in the specification. The	
CC	invention further comprises: an isolated polynucleotide encoding a	
CC	polypeptide with biological activity, where the polynucleotide hybridizes	
CC	to one of the 235 novel polynucleotides under stringent hybridization	
CC	conditions, or having greater than about 99% sequence identity with the	
CC	novel polynucleotide; a vector comprising a novel polynucleotide; an	
CC	expression vector comprising the novel polynucleotide; a host cell	
CC	genetically engineered to comprise the novel polynucleotide, which can be	
CC	operatively associated with a regulatory sequence that modulates	
CC	expression of the polynucleotide in the host cell; an isolated	
CC	polypeptide encoded by the novel polynucleotide, or a polynucleotide	
CC	hybridizing under stringent conditions to the novel polynucleotide; a	
CC	composition comprising the polypeptide and a carrier; an antibody	

CC directed against the polypeptide; a method for detecting the novel
CC polynucleotide in a sample; a method for detecting the polypeptide in a
CC sample; a method for identifying a compound that binds to the polypeptide
CC ; a method for producing the polypeptide; an isolated polypeptide
CC comprising any of the 235 amino acid sequences described in the
CC specification; and a collection of polynucleotides comprising of at least
CC one of the polynucleotides cited above. The polypeptides and
CC polynucleotides of the invention have antiinflammatory, cytostatic, and
CC antimicrobial activities. The novel polynucleotide may be used to treat
CC disorders by gene therapy. The polypeptides and polynucleotides are
CC useful for treating inflammation, leukaemias, nervous system disorders,
CC or infections. This sequence represents one of the 235 novel isolated
CC polynucleotides of the invention.

SQ Sequence 4213 BP; 702 A; 1368 C; 1320 G; 823 T; 0 U; 0 Other;

Query Match 97.4%; Score 751; DB 13; Length 4213;
Best Local Similarity 100.0%; Pred. No. 1e-157;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACCTCAAGAGCCTGAG 76
DB 3236 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACCTCAAGAGCCTGAG 3295
QY 77 CCAGCAGATCGAGAAACATCCGGAGCCCGAGAGGGAAGCCGCAAGAACCCCGCCGCACTG 136
DB 3296 CCAGCAGATCGAGAAACATCCGGAGCCCGAGAGGGAAGCCGCAAGAACCCCGCCGCACTG 3355
QY 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 196
DB 3356 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 3415
QY 197 CCAAGGCTGCAACTGTGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
DB 3416 CCAAGGCTGCAACTGTGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 3475
QY 257 CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAGACTGTACTCAGCAAGAACCCCAA 316
DB 3476 CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAGACTGTACTCAGCAAGAACCCCAA 3535
QY 317 GGAACAAGGAGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
DB 3536 GGAACAAGGAGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 3595
QY 377 CGGCCAGGGCTCCGACCTCCGATGCGCATCCAGCTGACCTTCTCGGCTGATGTC 436
DB 3596 CGGCCAGGGCTCCGACCTCCGATGCGCATCCAGCTGACCTTCTCGGCTGATGTC 3655
QY 437 CACGAGGCTCCAGAGATCATCTACCTACCTGCAAGACAGCGTGGCTACATGACCA 496
DB 3656 CACGAGGCTCCAGAGATCATCTACCTACCTGCAAGACAGCGTGGCTACATGACCA 3715
QY 497 GCAGACTGGCAACTCAAGAAGCCCTGCTCCCTCAAGGGCTCCCAAGAGATCAGATCCG 556
DB 3716 GCAGACTGGCAACTCAAGAAGCCCTGCTCCCTCAAGGGCTCCCAAGAGATCAGATCCG 3775
QY 557 CGCCGAGGGCAACAGCGGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCACAC 616
DB 3776 CGCCGAGGGCAACAGCGGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCACAC 3835
QY 617 CGAGGCTGGGGCAAGACAGTGTGTAATCAAAACACCAAGTCTCCCGCTGCCCAT 676
DB 3836 CGAGGCTGGGGCAAGACAGTGTGTAATCAAAACACCAAGTCTCCCGCTGCCCAT 3895
QY 677 CATCGATGTGGCCCTTGGAGCGTGTGGTGGCCAGACCAAGATTCGGCTTCGACGTTGG 736
DB 3896 CATCGATGTGGCCCTTGGAGCGTGTGGTGGCCAGACCAAGATTCGGCTTCGACGTTGG 3955
QY 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
DB 3956 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 3986

RESULT 10
ACCS0109
ID ACC50109 standard; cDNA; 5921 BP.
XX AC ACC50109;
DT 12-JUN-2003 (first entry)
XX Breast cancer associated cDNA sequence SEQ ID NO:66.
DE Human; breast cancer; cytostatic; gene therapy; gene; ss.
XX Homo sapiens.
XX WO2003004989-A2.
XX 16-JAN-2003.
XX 21-JUN-2002; 2002WO-US019669.
XX 21-JUN-2001; 2001US-0299887P.
PR 27-JUN-2001; 2001US-0301572P.
PR 18-JUL-2001; 2001US-0306501P.
PR 25-SEP-2001; 2001US-0325002P.
PR 05-MAR-2002; 2002US-0362585P.
PR 14-MAY-2002; 2002US-0380391P.
XX (MILL-) MILLENIUM PHARM INC.
XX Lillie J, Gannavarapu M, Glatt K, Hoerh S, Kamatkar S;
PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
PI Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
XX NPI; 2003-210381/20.
PT Breast cancer diagnosis or treatment by comparing the level of expression
PT of a marker in a patient sample with that in the control non-breast
PT cancer sample.
PS Claim 1; SEQ ID NO 66; 128pp; English.
XX The present invention describes a method for assessing whether a patient
CC is afflicted with breast cancer. The method comprises comparing the level
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC ABR47386 to ABR47632) in a patient sample and the normal level of
CC expression of the marker in a control non-breast cancer sample, where a
CC significant increase in the level of expression of the marker in the
CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences from
CC the present invention have cytostatic activities and can be used in gene
CC therapy. The method is useful for diagnosing and treating breast cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
Query Match 97.4%; Score 751; DB 8; Length 5921;
Best Local Similarity 100.0%; Pred. No. 1.1e-157;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACCTCAAGAGCCTGAG 76
DB 3776 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACCTCAAGAGCCTGAG 3835
QY 77 CCAGCAGATCGAGAAACATCCGGAGCCCGAGAGGGAAGCCCGCAAGAACCCCGCCGCACTG 136
DB 3836 CCAGCAGATCGAGAAACATCCGGAGCCCGAGAGGGAAGCCCGCAAGAACCCCGCCGCACTG 3895
QY 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 196
DB 3896 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 3955

QY 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 256
Db 3956 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 4015
QY 257 CGTGATCCCACTCAGCCCACTGGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA 316
Db 4016 CGTGATCCCACTCAGCCCACTGGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA 4075
QY 317 GGACAGAGGCGATGTCTGGTTCGGCGAGAGCATGACCGATGGATCCAGTTCCAGATATGG 376
Db 4076 GGACAGAGGCGATGTCTGGTTCGGCGAGAGCATGACCGATGGATCCAGTTCCAGATATGG 4135
QY 377 CGGCCAGGGCTCCGACCTGCCCATGTGGCCATCCAGCTGACCTTCTCGCGCTGATGTC 436
Db 4136 CGGCCAGGGCTCCGACCTGCCCATGTGGCCATCCAGCTGACCTTCTCGCGCTGATGTC 4195
QY 437 CACCGAGGCTCCAGAAATCACTACCACTGCAAGAAACAGCGTGGCTTACATGACCA 496
Db 4196 CACCGAGGCTCCAGAAATCACTACCACTGCAAGAAACAGCGTGGCTTACATGACCA 4255
QY 497 GCAGACTGGCAACCTCAGAGAGGCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556
Db 4256 GCAGACTGGCAACCTCAGAGAGGCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 4315
QY 557 CGCCGAGGGCAACAGCGCTTACCTACAGCGTCACTGTGATGGCTGCAACAGTCAACAC 616
Db 4316 CGCCGAGGGCAACAGCGCTTACCTACAGCGTCACTGTGATGGCTGCAACAGTCAACAC 4375
QY 617 CGAGGCTGGGGCAAGACAGTGAATGAATCAAAACACCAAGTCTCCGCTGCGCCAT 676
Db 4376 CGAGGCTGGGGCAAGACAGTGAATGAATCAAAACACCAAGTCTCCGCTGCGCCAT 4435
QY 677 CATCGATGGCCCTTGGACCTTGGTGGCCAGACCGAGATTCGGCTTCGACCTGG 736
Db 4436 CATCGATGGCCCTTGGACCTTGGTGGCCAGACCGAGATTCGGCTTCGACCTGG 4495
QY 737 CCCTGTCTGCTTCTGTAACCTCCCTCCATC 767
Db 4496 CCCTGTCTGCTTCTGTAACCTCCCTCCATC 4526

RESULT 11
ID ADP65048
XX ADP65048 standard; DNA; 5921 BP.
XX ADP65048;
XX 12-AUG-2004 (first entry)
XX Human collagen, type I, alpha 1 (COL1A1) DNA sequence.
XX autoimmune disease; arthritis; gene expression analysis;
XX rheumatoid arthritis; collagen-induced; immunosuppressive; antirheumatic;
XX antiarthritic; osteopathic; antigout; antiinflammatory; dermatological;
XX immunomodulatory; lupus; ankylosing spondylitis; Fibrositis;
XX fibromyalgia; Osteoarthritis; gout; juvenile rheumatoid arthritis;
XX immune; ds; human.
XX Homo sapiens.
XX OS
XX PN WO2003072827-A1.
XX PD 04-SEP-2003.
XX PF 31-OCT-2002; 2002WO-US035433.
XX PR 31-OCT-2001; 2001US-0336220P.
XX PA (CHIL-) CHILDREN'S HOSPITAL MEDICAL CENT.
XX Hirsch R, Thorton SL;
XX WPI; 2003-712740/67.
Db

DR GENBANK; NM_000088.
XX Diagnosing and analyzing autoimmune disease using gene expression
PT profiles and microarray technology, useful for diagnosing and treating
PT rheumatoid arthritis, lupus, fibrositis, osteoarthritis, fibromyalgia and
PT gout.
XX Disclosure; Page; 56pp; English.
XX The invention relates to a novel method for diagnosing and analysing
CC autoimmune disease or arthritides. The method comprises obtaining a
CC patient sample containing mRNA, analysing gene expression using the mRNA
CC that results in a gene expression signature of the mRNA, and using that
CC gene expression signature to diagnose or analyse the autoimmune disease
CC or arthritides in the patient, where gene expression of at least 60% of
CC the genes correlates with that of the gene signature. The invention
CC further comprises: a treatment of rheumatoid arthritis; identification of
CC genes for targeting in the treatment of rheumatoid arthritis in a mammal
CC other than a mouse; diagnosis of rheumatoid arthritis in a mammal; an
CC array or gene chip, specific for rheumatoid arthritis; diagnosis or
CC analyses of autoimmune disease or rheumatoid arthritis; screening the
CC efficacy of a candidate drug in vitro for the treatment of collagen-
CC induced arthritis; and reducing the symptoms associated with collagen-
CC induced arthritis. The compositions of the invention have the following
CC activities: immunosuppressive, antirheumatic, antiarthritic, osteopathic,
CC antigout, antiinflammatory, dermatological, and immunomodulatory. The
CC methods and compositions of the present invention are useful for
CC diagnosing and treating autoimmune disease or arthritides, such as
CC rheumatoid arthritis, lupus, ankylosing spondylitis, fibrositis,
CC fibromyalgia, osteoarthritis, gout, juvenile rheumatoid arthritis, and an
CC immune disease caused by an infectious agent. This polynucleotide
CC represents a DNA sequence relating to the genes used in the analysis and
CC treatment of autoimmune diseases or arthritides. Note: This sequence is
CC not shown in the specification. It has been supplied in an electronic
CC format from WIPO.
XX

QY Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
XX
XX Query Match 97.4%; Score 751; DB 11; Length 5921;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-157;
XX Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 TGATGCCAATGTGGTTCTGTCACCTCGAGGTGGACACCACTCAAGAGCTGAG 76
Db 3776 TGATGCCAATGTGGTTCTGTCACCTCGAGGTGGACACCACTCAAGAGCTGAG 3835
QY 77 CCAGCAGATCGAGAACATCCGGAGCCCGAGGGAGCCGACAGACCCCGCCGACCTG 136
Db 3836 CCAGCAGATCGAGAACATCCGGAGCCCGAGGGAGCCGACAGACCCCGCCGACCTG 3895
QY 137 CGGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTCTGGATTGACCCCAA 196
Db 3896 CGGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTCTGGATTGACCCCAA 3955
QY 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 256
Db 3956 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 4015
QY 257 CGTGATCCCACTCAGCCCACTGGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA 316
Db 4016 CGTGATCCCACTCAGCCCACTGGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA 4075
QY 317 GGACAGAGGCGATGTCTGGTTCGGCGAGAGCATGACCGATGGATCCAGTTCCAGATATGG 376
Db 4076 GGACAGAGGCGATGTCTGGTTCGGCGAGAGCATGACCGATGGATCCAGTTCCAGATATGG 4135
QY 377 CGGCCAGGGCTCCGACCTGCCCATGTGGCCATCCAGCTGACCTTCTCGCGCTGATGTC 436
Db 4136 CGGCCAGGGCTCCGACCTGCCCATGTGGCCATCCAGCTGACCTTCTCGCGCTGATGTC 4195
QY 437 CACCGAGGCTCCAGAAATCACTACCACTGCAAGAAACAGCGTGGCTTACATGACCA 496
Db 4196 CACCGAGGCTCCAGAAATCACTACCACTGCAAGAAACAGCGTGGCTTACATGACCA 4955

QY 497 GCAGATGGCAACCTCAAGAGGCGCTCTCTCAAGGGCTCCCAAGAGATCGAGATCCG 556
 DB |||||
 DB 4256 GCAGATGGCAACCTCAAGAGGCGCTCTCTCAAGGGCTCCCAAGAGATCGAGATCCG 4315
 QY 557 GCGAGAGGCAACAGCCGCTTCACTACAGCGTCACTGTTCGATGGCTGCAGAGTCAAC 616
 DB |||||
 DB 4316 GCGAGAGGCAACAGCCGCTTCACTACAGCGTCACTGTTCGATGGCTGCAGAGTCAAC 4375
 QY 617 CGGAGCTGGGGCAAGACAGTGAATACAAACCAACCAAGTCTCCCGCTGCCCAT 676
 DB |||||
 DB 4376 CGGAGCTGGGGCAAGACAGTGAATACAAACCAACCAAGTCTCCCGCTGCCCAT 4435
 QY 677 CATGATGTGGCCCGCTTGGACGTTGGTGGCCCAAGACAGCAATTCGAGCTTGG 736
 DB |||||
 DB 4436 CATGATGTGGCCCGCTTGGACGTTGGTGGCCCAAGACAGCAATTCGAGCTTGG 4495
 QY 737 CCTGTCTGCTTCTGTAAACTCCCTCCATC 767
 DB |||||
 DB 4496 CCTGTCTGCTTCTGTAAACTCCCTCCATC 4526

RESULT 12

ADQ19469
 ID ADQ19469 standard; DNA; 5921 BP.

XX AC ADQ19469;

XX XX 26-AUG-2004 (first entry)

XX XX Human soft tissue sarcoma-upregulated DNA - SEQ ID 2288.

XX XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human;
 XX KW ds.

XX OS Homo sapiens.

XX XX W02004048938-A2.

XX XX 10-JUN-2004.

XX XX 26-NOV-2003; 2003WO-US038193.

XX XX 26-NOV-2002; 2002US-0429739P.

XX XX (PROT-) PROTEIN DESIGN LABS INC.

XX XX Aziz N, Ginsburg WM, Zlotnik A;

XX XX WPI; 2004-441208/41.

XX XX Early detection of soft tissue sarcoma comprises determining expression
 PT of a gene in a first soft tissue sample and a normal soft tissue sample
 PT and comparing the gene expression, also useful in treating soft tissue
 PT sarcoma.

XX XX Example 2; SEQ ID NO 2288; 210pp; English.

XX XX The invention relates to a novel method for detecting soft tissue sarcoma
 CC which comprises obtaining a first soft tissue sample from an individual
 CC and a normal soft tissue sample from the same or different individual,
 CC determining the expression of a gene in both samples and comparing the
 CC expression of the gene in both soft tissue samples, where a higher level
 CC of protein expression in the first soft tissue sample indicates the
 CC presence of soft tissue sarcoma. The method of the invention has
 CC cytostatic applications and may be useful for detecting soft tissue
 CC sarcoma, possibly via gene therapy or vaccine production. The nucleic
 CC acid sequences may be useful in diagnostic and screening applications.
 CC The current sequence is that of a human soft tissue sarcoma-upregulated
 CC DNA of the invention. The current sequence is not shown within the
 CC specification per se but was submitted in CD format by the inventor.

XX XX Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;

Query Match 97.4%; Score 751; DB 12; Length 5921;
 Best Local Similarity 100.0%; Pred. No. 1.1e-157; Indels 0; Gaps 0;
 Matches 751; Conservative 0; Mismatches 0;

QY 17 TGATGCCAATGTGGTTCTGTGACCGTGCACCTCGAGGTGGACACCACTCTCAAGAGCTTGAG 76
 DB |||||
 DB 3776 TGATGCCAATGTGGTTCTGTGACCGTGCACCTCGAGGTGGACACCACTCTCAAGAGCTTGAG 3835
 QY 77 CCAGCAGATCGAGAAATATCCGAGCCCGAGAGGGAAGCCGCAAGAACCCCGCCCGACCTG 136
 DB |||||
 DB 3836 CCAGCAGATCGAGAAATATCCGAGCCCGAGAGGGAAGCCGCAAGAACCCCGCCCGACCTG 3895
 QY 137 CCGTGACCTCAAGATGTGCCACTCTGACTGAGAGTGGAGAGTACTGGATTGAGCCCAA 196
 DB |||||
 DB 3896 CCGTGACCTCAAGATGTGCCACTCTGACTGAGAGTGGAGAGTACTGGATTGAGCCCAA 3955
 QY 197 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGTGTGAGACCTG 256
 DB |||||
 DB 3956 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGTGTGAGACCTG 4015
 QY 257 CGTGATCCCACTCAGCCCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
 DB |||||
 DB 4016 CGTGATCCCACTCAGCCCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 4075
 QY 317 GGACAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
 DB |||||
 DB 4076 GGACAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 4135
 QY 377 CGGCCAGGGCTCCGACCCCTCGATGTGGCCATCCAGTCCAGTCCCTTCTGCGCCCTGATGTC 436
 DB |||||
 DB 4136 CGGCCAGGGCTCCGACCCCTCGATGTGGCCATCCAGTCCAGTCCCTTCTGCGCCCTGATGTC 4195
 QY 437 CACCGAGGCTCCCGAGAAATCACTACCACTGCAAGAACAGCGTGGCCTACATGGACCA 496
 DB |||||
 DB 4196 CACCGAGGCTCCCGAGAAATCACTACCACTGCAAGAACAGCGTGGCCTACATGGACCA 4255
 QY 497 GCAGATGGCAACCTCAAGAGGCGCTCTCTCAAGGGCTCCCAAGAGATCGAGATCCG 556
 DB |||||
 DB 4256 GCAGATGGCAACCTCAAGAGGCGCTCTCTCAAGGGCTCCCAAGAGATCGAGATCCG 4315
 QY 557 GCGCAGGGCAACAGCCGCTTCACTACAGCGTCACTGTTCGATGGCTGCAGACTCAAC 616
 DB |||||
 DB 4316 GCGCAGGGCAACAGCCGCTTCACTACAGCGTCACTGTTCGATGGCTGCAGACTCAAC 4375
 QY 617 CGGAGCTGGGGCAAGACAGTGAATACAAACCAACCAAGTCTCTCCCGCTGCCCAT 676
 DB |||||
 DB 4376 CGGAGCTGGGGCAAGACAGTGAATACAAACCAACCAAGTCTCTCCCGCTGCCCAT 4435
 QY 677 CATCGATGTGGCCCGCTTGGACGTTGGTGGCCCGAGACCAAGAAATTCGGCTTCGACGTTGG 736
 DB |||||
 DB 4436 CATCGATGTGGCCCGCTTGGACGTTGGTGGCCCGAGACCAAGAAATTCGGCTTCGACGTTGG 4495
 QY 737 CCTGTCTGCTTCTGTAAACTCCCTCCATC 767
 DB |||||
 DB 4496 CCTGTCTGCTTCTGTAAACTCCCTCCATC 4526

RESULT 13

ADQ29586
 ID ADQ29586 standard; DNA; 5921 BP.

XX AC ADQ29586;

XX XX 07-OCT-2004 (first entry)

XX XX Human colorectal cancer-associated protein coding sequence #8.

XX XX human; colon cancer; TIMP1; RegI-alpha;

XX KW colorectal cancer-associated marker; gene; ds.

XX XX Homo sapiens.

XX XX

CC	an anti-cancer agent. M4 is useful for identifying a compound for
CC	treating breast cancer. Note: The sequence data for this patent did not
CC	form part of the printed specification, but was obtained in electronic
XX	format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
QQ	
SQ	Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
	Query Match 97.4%; Score 751; DB 13; Length 5921;
	Best Local Similarity 100.0%; Pred. No. 1.le-157;
	Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	17 TGATGCCAATGTGTTGGTACCGTGACCTTCAGAGTGGAACCAACCCTCAAGAGCCTGAG 76
Dd	3776 TGATGSCAAATGTGTTGTGACCGTGACCTTCAGAGTGGAACCAACCCTCAAGAGCCTGAG 3835
Qy	77 CCACGACATCGAGAACAATCCGGAGCCACAGAGGGAGCCGCAAGAAACCCC GCCGCACCTG 136
Dd	3836 CCACGACATCGAGAACAATCCGGAGCCACAGAGGGAGCCGCAAGAAACCCC GCCGCACCTG 3895
Qy	137 CCGTGACTCAAAGATGTGCCACTCTGACTGGAAGAGTGGAGGTAFTGTGGATTGACCCCAA 196
Dd	3896 CCGTGACTCAAAGATGTGCCACTCTGACTGGAAGAGTGGAGGTAFTGTGGATTGACCCCAA 3955
Qy	197 CCAAGGCTGC AA ACTGGATGCCATCA A AGTCTTCTGCAACATGGAGACTGGTGGAGACTG 256
Dd	3956 CCAAGGCTGC AA ACTGGATGCCATCA A AGTCTTCTGCAACATGGAGACTGGTGGAGACTG 4015
Qy	257 CGTGTAACCCACTCAGGCCGATGTGGGCCAGAGAATCTGTATCATGACGAAGAACCCCAA 316
Dd	4016 CGTGTAACCCACTCAGGCCGATGTGGGCCAGAGAATCTGTGTATCATGACGAAGAACCCCAA 4075
Qy	317 GGACAAGAGGATCTCTGTGTTTCGGCGAGAGCATCACCGATGGATTCCAGTTCCAGATATGG 376
Dd	4076 GGA CAAGAGGATCTCTGTGTTTCGGCGAGAGCATCACCGATGGATTCCAGTTCCAGATATGG 4135
Qy	377 CGGCCAGGGCTCCAGACCTCGCCGATGTGGCCATCCAGCTGACCTTCTCTCGCCCTGATGTC 436
Dd	4136 CGGCCAGGGCTCCAGACCTCGCCGATGTGGCCATCCAGCTGACCTTCTCTCGCCCTGATGTC 4195
Qy	437 CACGAGGCTCC CAGAACATCACTTACCAATGCAAGAACAGCGTGCCCTACATGGACCA 496
Dd	4196 CACGAGGCTCC CAGAACATCACTTACCAATGCAAGAACAGCGTGCCCTACATGGACCA 4255
Qy	497 GCAGACTGGCAACTCAAGAGGGCCCTGCTCCTCAAGGGCTCCAACAGAGATCGAGATCCG 556
Dd	4256 GCAGACTGGCAACTCAAGAGGGCCCTGCTCCTCAAGGGCTCCAACAGAGATCGAGATCCG 4315
Qy	557 CGCCGAGGGCAACAGCGCTTCACTTACAGGTCACTGTGATGGCTGCAAGAGTCAACAC 616
Dd	4316 CGCCGAGGGCAACAGCGCTTCACTTACAGGTCACTGTGATGGCTGCAAGAGTCAACAC 4375
Qy	617 CGGAGCTGGGGCAAGACAGTGA TTGAATAACAAAACCAACGATCTCTCCCGCTGCCCAT 676
Dd	4376 CGGAGCTGGGGCAAGACAGTGA TTGAATAACAAAACCAACGATCTCTCCCGCTGCCCAT 4435
Qy	677 CATCGATGTGCCCCCTTGGACGTTGTGGTGGCCCCACAGCAGGAATTCGGCTTCGACGTTGG 736
Dd	4436 CATCGATGTGCCCCCTTGGAGGTTGTGTGCCCCACAGCAGGAATTCGGCTTCGACGTTGG 4495
Qy	737 CCTGTCTGTTCTCTGTA AACCTCCCTCCATC 767
Dd	4496 CCTGTCTGTTCTCTGTA AACCTCCCTCCATC 4526

RESULT 15	
ADZ26572	
ID	ADZ26572 standard; cDNA; 5921 BP.
XX	
XX	
AC	ADZ26572;
XX	
DT	16-JUN-2005 (first entry)
XX	
XX	
DE	Human type I collagen cDNA.

XX	cell culture; stem cell; type I collagen; ss; gene.
KW	Homo sapiens.
XX	WO2005030999-A1.
OS	07-APR-2005.
PN	24-SEP-2004; 2004WO-US031524.
XX	25-SEP-2003; 2003US-0506221P.
XX	08-OCT-2003; 2003US-0509594P.
PD	(DAND) DANA FARBER CANCER INST INC.
XX	Ritz J, Wu CJ;
PI	WPI; 2005-273394/28.
XX	P-PSDB; ADZ26573.
DR	Detecting lineage-specific cells in a biological sample, useful for
DR	determining the clinical outcome of a progenitor cell transfer in a
XX	subject, comprises identifying lineage-specific mRNA in the sample.
PT	Disclosure; SEQ ID NO 198; 393pp; English.
XX	The invention relates to a method of detecting lineage-specific cells in
CC	a biological sample which comprises identifying lineage-specific mRNA in
CC	the sample. The methods are useful for determining the clinical outcome
CC	of a progenitor cell transfer in a subject, and for identifying or
CC	quantifying lineage-specific cells. The present sequence represents a
CC	human cDNA that encodes a protein used to identify lineage-specific
CC	cells.
XX	Sequence 5921 BP; 1068 A; 1879 C; 1763 G; 1211 T; 0 U; 0 Other;
XX	Query Match 97.4%; Score 751; DB 14; Length 5921;
XX	Best Local Similarity 100.0%; Pred. No. 1.1e-157;
XX	Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	17 TGATGCCAAATGTGGTTCGTGCACCGTGCACCTCGAGGTGGACACCACTCAAGAGCCTGAG 76
Db	3776 TGATGCCAAATGTGGTTCGTGCACCGTGCACCTCGAGGTGGACACCACTCAAGAGCCTGAG 3835
QY	77 CCAGCAGATCGAACAATCCGGAGCCCGAGAGGAAAGCCGCAAGAACCCGCCGCACTG 136
Db	3836 CCAGCAGATCGAACAATCCGGAGCCCGAGAGGAAAGCCGCAAGAACCCGCCGCACTG 3895
QY	137 CCGTGACCTCAAGTGTGCCACTCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 196
Db	3896 CCGTGACCTCAAGTGTGCCACTCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 3955
QY	197 CCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCACAATCGAGACTGGTCAGACTG 256
Db	3956 CCAAGGCTGCAACCTGGATGCCATCAAAGTCTTCTGCACAATCGAGACTGGTCAGACTG 4015
QY	257 CGTGTAACCCCACTCAGCCCAAGTGTGGCCCAAGAGAACTGTGTAATCAGCAAGAACCCCAA 316
Db	4016 CGTGTAACCCCACTCAGCCCAAGTGTGGCCCAAGAGAACTGTGTAATCAGCAAGAACCCCAA 4075
QY	317 GGCACAGAGGCATGTCGTGGTTCGGCGAGAGCATGACCGATGGATTCAGATTCCAGTATGG 376
Db	4076 GGCACAGAGGCATGTCGTGGTTCGGCGAGAGCATGACCGATGGATTCAGATTCCAGTATGG 4135
QY	377 CGGCCAGGGCTCCGACCTTCGCCGATGTGGCCATTCAGCTGACCTTCTCTGGCCCTGATGTC 436
Db	4136 CGGCCAGGGCTCCGACCTTCGCCGATGTGGCCATTCAGCTGACCTTCTCTGGCCCTGATGTC 4195
QY	437 CACCGAGGCTCCCGAAGAACTACCTTACCTGCAAGAAACAGCGTGGCTTACATGACCA 496
Db	4196 CACCGAGGCTCCCGAAGAACTACCTTACCTGCAAGAAACAGCGTGGCTTACATGACCA 4255

Qy	497	GCAGACTGGCAACCTCAAGAAGCCCTGTCTCTCAAGGGCTCCAACGAGATCGAGATCCG	556
Db	4256	GCAGACTGGCAACCTCAAGAAGCCCTGTCTCTCAAGGGCTCCAACGAGATCGAGATCCG	4315
Qy	557	CGCCGAGGGCAACAGCGGCTTCACTACAGCGTCACTGTGATGGCTGCACGAGTCACAC	616
Db	4316	CGCCGAGGGCAACAGCGGCTTCACTACAGCGTCACTGTGATGGCTGCACGAGTCACAC	4375
Qy	617	CGGAGCTGGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGTCTCCCGCTGCCCAT	676
Db	4376	CGGAGCTGGGGCAAGACAGTGAATTGAATACAAAACCAACCAAGTCTCCCGCTGCCCAT	4435
Qy	677	CATCGATGTGGCCCCCTTGGACGTTGGTGCCCCCAGACACAGGAATTCGGGCTTCGACGTTGG	736
Db	4436	CATCGATGTGGCCCCCTTGGACGTTGGTGCCCCCAGACACAGGAATTCGGGCTTCGACGTTGG	4495
Qy	737	CCCTGTCTGCTTCTGTAAACTCCCTCCATC	767
Db	4496	CCCTGTCTGCTTCTGTAAACTCCCTCCATC	4526

Search completed: April 27, 2006, 19:33:57
Job time : 484.986 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 17:29:04 ; Search time 3924.36 Seconds
(without alignments)
9192.024 Million cell updates/sec

Title: US-10-677-877A-3
Perfect score: 771
Sequence: 1 aagcttacgtaagatctgat.....gtaaacctccctccatctaga 771

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_gse1.*
10: gb_gse2.*
11: gb_gse3.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	734.2	95.2	3492	10 AY414729	AY414729 Homo sapi
2	715.6	92.8	762	7 CN342438	CN342438 170004250
3	699.8	90.8	904	8 DR768506	DR768506 ILLUMIGEN
C 4	697	90.4	1035	1 AL568796	AL568796 AL568796
5	684	88.7	867	8 DN103076	DN103076 1098161 M
6	680.6	88.3	694	7 CN342173	CN342173 170005999
7	680	88.2	820	3 BP172484	BP172484 BP172484
8	680	88.2	3378	10 AY414730	AY414730 Pan trogl
9	674.8	87.5	874	3 BP172845	BP172845 BP172845
10	674.2	87.4	700	7 CN342548	CN342548 170004249
11	673.6	87.4	813	7 CV547993	CV547993 EST409 Bo
12	673	87.3	800	3 BP172993	BP172993 BP172993
C 13	672	87.2	796	6 CD607437	CD607437 56095953J
14	671.2	87.1	701	7 CN342240	CN342240 170005321
C 15	671	87.0	696	8 DR432524	DR432524 AGENCOURT
16	669.4	86.8	791	7 CU023328	CU023328 CU023328
C 17	664.8	86.2	682	8 DR432984	DR432984 AGENCOURT
18	661	85.7	721	1 AW577435	AW577435 QV3-BT053
19	660.6	85.7	667	7 CN394438	CN394438 170004245
C 20	658	85.3	671	8 DR433321	DR433321 AGENCOURT
C 21	655.4	85.0	674	7 CN342343	CN342343 170004245
22	654.6	84.9	675	7 CN342526	CN342526 170005313

C	23	651.6	84.5	764	6	CD607434
	24	649.6	84.3	683	7	CN342365
	25	647.2	83.9	652	7	CN342584
C	26	642.8	83.4	755	6	CD607387
	27	642.2	83.3	789	8	DN103391
C	28	642.2	83.3	1275	6	CF111062
	29	642.2	83.3	1275	6	CF111063
	30	641.6	83.2	1088	2	BG113776
	31	641.2	83.2	778	3	BP172582
	32	640.6	83.1	703	7	CV674726
	33	638.2	82.8	663	7	CN342398
	34	634.4	82.3	656	3	BI916706
	35	634.2	82.3	639	7	CN342371
	36	631.4	81.9	890	6	CA976611
	37	630.2	81.7	729	6	CD607386
	38	628.2	81.5	868	6	CF618321
	39	627.6	81.4	700	6	CF765818
	40	625.8	81.2	637	2	BG938755
	41	625.6	81.1	734	2	BI181563
	42	625	81.1	719	2	BI184997
	43	625	81.1	721	2	BI181198
	44	624.2	81.0	769	6	CD607436
45	622.4	80.7	874	6	CB589398	

ALIGNMENTS

RESULT 1
AY414729 3492 bp DNA linear GSS 17-DEC-2003
LOCUS Homo sapiens COL1A1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY414729
VERSION AY414729.1 GI:39770691
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 3492)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 3492)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
Todd,M.A., Tanenbaum,D.M., Civello,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source Location/Qualifiers
1..3492
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
gene <1..>3492
/locus_tag="HCM5321"
ORIGIN
Query Match 95.2%; Score 734.2; DB 10; Length 3492;
Best Local Similarity 99.8%; Pred.No. 2.3e-165;
Matches 736; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Db	701	TGTGGCCCCCTTGACGTTGGTGGCCCCCAGACCAGGAATTCGGCTTCGACGTTGGCCCTGT	760
Qy	743	CT 744	
Db	761	CT 762	
RESULT 3			
DR768506			
LOCUS			
DEFINITION			
ILLUMINIGEN MQ_61545 Katze MMOV Macaca mulatta cDNA clone			
IBIUM:31396 57 similar to Bases 5 to 804 highly similar to human			
COL1A1 (Hs.172928), mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
COMMENT			
FEATURES			
source			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
Qy	17	TGATGCCAATGTGTTCTGTCACCGTGCCTCGAGTGGACACACACCTCAAGAGCCTGAG	76
Db	44	TGATGCCAATGTGTTCTGTCACCGTGCCTCGAGTGGACACACACCTCAAGAGCCTGAG	103
Qy	77	CCAGCAGATCGAGACATCCGAGGCCAGAGGGGAGCCGCAAGAACCCCGCCGACCTG	136
Db	104	CCAGCAGATCGAGAACATCCGAGGCCAGAGGGGAGCCGCAAGAACCCCGCCGACCTG	163
Qy	137	CCGTGACCTCAAGATGTCACACTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA	196
Db	164	CCCGACCTCAAGATGTGCCACTCTGACTGGAAAGAGCGGAGATCTGGATTGACCCCAA	223
Qy	197	CCAAGGCTGCAACCTCGATGCCATCAAAAGTCTTCTGCAACATCGAGACTGGTGAGACCTG	256
Db	224	CCAAGGCTGCAACCTCGATGCCATCAAAAGTCTTCTGCAACATCGAGACTGGTGAGACCTG	283
Qy	257	CGTGTATCCCACTCAGCCCAAGTGTGGCCCAAGAAACTTGTTATCATCAGCAAGAACCCCAA	316
Db	284	CGTGTATCCCACTCAGCCCAAGTGTGGCCCAAGAAACTTGTTATCATCAGCAAGAACCCCAA	343
Qy	317	GGACAAGGCACTGTCTGGTTGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	376
Db	344	GGACAAGGCACTGTCTGGTTGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	403
Qy	377	CGCCCAAGGCTCCGACCTCGCCGATGTGGCCATTCAGCTGACCTTCTCGCGCTGATGTC	436
Db	404	CGCCCAAGGCTCCGACCTCGCCGATGTGGCCATTCAGCTGACCTTCTCGCGCTGATGTC	463
Qy	437	CACCGAGGCTCCCAAGAACTACCTACCACTGCAAGAAACAGCGTGGCCTTACATGAGCA	496
Db	464	CACCGAGGCTCCCAAGAACTACCTACCACTGCAAGAAACAGCGTGGCCTTACATGAGCA	523
Qy	497	GCAGATGGCAACCTCAAGAGGCTTCTCTCAAGGGCTCCCAAGAGATCGAGATCCG	556
Db	524	GCAGATGGCAACCTCAAGAGGCTTCTCTCAAGGGCTCCCAAGAGATCGAGATCCG	583
Qy	557	CGCCGAGGCAACAGCGCTTACCTACAGCGTCACTGTCTGATGGCTGCACAGATCACAC	616
Db	584	CGCCGAGGCAACAGCGTCACTGTCTGATGGCTGCACAGATCACAC	643
Qy	617	CGAGCCTGGGCAAGACAGTGAATGAATAACCAACACCAAGTCTCTCCGCTGCGCAT	676
Db	644	CGAGCCTGGGCAAGACAGTGAATGAATAACCAACACCAAGTCTCTCCGCTGCGCAT	702
Qy	677	CATCGATGGGCCCCCTTGGAGTTGGTGGCCCAAGACAGGATTCGGCTTCACGTTGG	736
Db	703	CATCGAGTGGCCCCCTTGGAGTTGGTGGCCCAAGACAGGATTCGGCTTCACGTTGG	762
Qy	737	CCCTGCTCTGCTCTGTAACCTCCCTCCATCTAGA	771
Db	763	CCCGTCTGGTCTTGGAACTCCCTCTCTCCACA	797
RESULT 4			
AL568796/c			
LOCUS			
DEFINITION			
AL568796 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE005YLL1			
3-PRIME, mRNA sequence.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
Qy	17	TGATGCCAATGTGTTCTGTCACCGTGCCTCGAGTGGACACACACCTCAAGAGCCTGAG	76
Db	44	TGATGCCAATGTGTTCTGTCACCGTGCCTCGAGTGGACACACACCTCAAGAGCCTGAG	103
Qy	77	CCAGCAGATCGAGACATCCGAGGCCAGAGGGGAGCCGCAAGAACCCCGCCGACCTG	136
Db	104	CCAGCAGATCGAGAACATCCGAGGCCAGAGGGGAGCCGCAAGAACCCCGCCGACCTG	163


```
Db 439 CGGCGAGGCTCCGATCTCTGTGACGTGGCCCATCCAGCTGACCTTCTCTGCGCCTGATGTC 498
Qy 437 CACCGAGGCTCCAGAACATACCTACCACTGCAAGAACAGAGCGTGGCCTACATGGACCA 496
Db 499 CACTGAGGCTTCCAGAACATACCTACCACTGCAAGAACAGAGCGTGGCCTACATGGACCA 558
Qy 497 GCAGACTGGCACTCAAGAGGCTGCTCTCTCAAGGCTCCAAACGAGATCGAGATCCG 556
Db 559 GCAGACTGGCACTCAAGAGGCTGCTCTCTCAAGGCTCCAAACGAGATCGAGATCCG 618
Qy 557 CGCGAGGCTCAAGAGGCTGCTCTCTCAAGGCTCCAAACGAGATCGAGATCCG 616
Db 619 GCGCGAGGCTCAAGAGGCTGCTCTCTCAAGGCTCCAAACGAGATCGAGATCCG 678
Qy 617 CGGAGGCTGGGCGCAAGACAGTGAATGAATCAAAACCAACCAAGCTTCCCGCCTGGCCAT 676
Db 679 CGGAGGCTGGGCGCAAGACAGTGAATGAATCAAAACCAACCAAGCTTCCCGCCTGGCCAT 738
Qy 677 CATGATGTGCCCCCTTGGAGCTTGGTGGCCCGACAGACAGGAATTCGGCTTCGAGCTGG 736
Db 739 CATGATGTGCCCCCTTGGAGCTTGGTGGCCCGACAGGAATTCGGCTTCGAGCTGG 798
Qy 737 CCCTGTCTGCTTCTGTAACTCCCTCC 764
Db 799 CCCTGTCTGCTTCTGTAACTCCCTCC 826

RESULT 6
CN342173 694 bp mRNA linear EST 16-MAY-2004
LOCUS 1700599934677 GRN_PRENH Homo sapiens cDNA 5', mRNA sequence.
DEFINITION CN342173
ACCESSION CN342173.1 GI:47342107
VERSION EST.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 694)
AUTHORS Brandenberger, R., Wei, H., Zhang S., Lei, S., Murage, J., Fisk, G.J.,
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
Lebkowski, J. and Stanton, L.W.
TITLE Transcriptional characterization elucidates signaling networks that
control human ES cell growth and differentiation
JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
PUBMED 15146197
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 694 Std Error: 0.00.
FEATURES
source
1..694
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cell, retinoic acid and
mitogen-treated hES cell line H7"
/clone_lib="GRN PRENEU"
/note="Oligo dT primed, full-length enriched cDNA library
from hES cell line H7 (p29) maintained in feeder-free
conditions. Embryoid bodies were generated in the presence
of all-trans retinoic acid and mitogens."

ORIGIN
Query Match 88.3%; Score 680.6; DB 7; Length 694;
Best Local Similarity 99.4%; Pred. No. 1.4e-152;
Matches 683; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
Qy 77 CCAGCAGATCGAGAACATCCGAGCCCGCAGGAGGAGCCGCAAGAACCCCGCCGACCTG 136
Db 8 CGAGCAGATCGAGAACATCCGAGCCCGCAGGAGGAGCCGCAAGAACCCCGCCGACCTG 67
Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA 196
Db 68 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA 127
Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGACTG 256
Db 128 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGACTG 187
Qy 257 CGTGTACCCCACTCAGCCAGTGTGGCCCGCAGAGAACTGCTCATCAGCAGAACCCCAA 316
Db 188 CGTGTACCCCACTCAGCCAGTGTGGCCCGCAGAGAACTGCTCATCAGCAGAACCCCAA 247
Qy 317 GGAACAAGAGGCTGTCTGGTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
Db 248 GGAACAAGAGGCTGTCTGGTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 307
Qy 377 CGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 436
Db 308 CGGCCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 367
Qy 437 CACCGAGGCTCCCGAGAACATCACTACATCACTGCAAGAACAGAGTGGCCTACATGGACCA 496
Db 368 CACCGAGGCTCCCGAGAACATCACTACATCACTGCAAGAACAGAGTGGCCTACATGGACCA 427
Qy 497 GCAGACTGGCACTCAAGAGGCTGCTCTCAAGGCTCCAAACGAGATCGAGATCCG 556
Db 428 GCAGACTGGCACTCAAGAGGCTGCTCTCAAGGCTCCAAACGAGATCGAGATCCG 487
Qy 557 CGCGAGGCGCAACAGCCGCTTCACTACAGCGTCACCTGTCGATGGCTGCACGAGTCACAC 616
Db 488 CGCGAGGCGCAACAGCCGCTTCACTACAGCGTCACCTGTCGATGGCTGCACGAGTCACAC 547
Qy 617 CGGAGCCTGGGCGCAAGACAGTGAATGAATCAAAACCAACCAAGTCTCCCGCCTGCCCAT 676
Db 548 CGGAGCCTGGGCGCAAGACAGTGAATGAATCAAAACCAACCAAGTCTCCCGCCTGCCCAT 607
Qy 677 CATCGATGTGGCCCTTGGAGCTTGGTGGCCCGACAGACAGGAATTCGGCTTCGAGCTGG 736
Db 608 CATCGATGTGGCCCTTGGAGCTTGGTGGCCCGACAGACAGGAATTCGGCTTCGAGCTGG 667
Qy 737 CCCTGTCTGCTTCTGTAACTCCCTCC 763
Db 668 CCCTGTCTGCTTCTGTAACTCCCTCC 694

RESULT 7
BP172484 820 bp mRNA linear EST 30-DEC-2003
LOCUS BP172484
DEFINITION BP172484 full-length enriched swine cDNA library, adult uterus Sus
scrofa cDNA clone UTR01022F11 5', mRNA sequence.
ACCESSION BP172484
VERSION BP172484.1 GI:40421957
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
REFERENCE 1 (bases 1 to 820)
AUTHORS Uenishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H.,
Okumura, N., Hamasima, N. and Awata, T.
TITLE PDE (Pig EST Data Explorer): construction of a database for ESTs
derived from porcine full-length cDNA libraries
JOURNAL Nucleic Acids Res. 32 (1), D484-D488 (2004)
PUBMED 14681463
COMMENT Contact: Hirohide Uenishi
Animal Genome Laboratory, Genome Research Department
National Institute of Agrobiological Sciences
```

2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
Tel: +81-29-838-8627
Fax: +81-29-838-8627
Email: huenish@aifrc.go.jp
EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA
library
Vector sequences were eliminated by RepeatMasker version 2002/07/13
and crossmatch version 0.990319
Low quality bases were trimmed based on the quality values.
FEATURES
source
1..820
/organism="Sus scrofa"
/mol_type="mRNA"
/db_xref="taxon:9823"
/clone="UTR010022F11"
/tissue_type="uterus"
/dev_stage="adult"
/clone_lib="full-length enriched swine cDNA library, adult
uterus"
ORIGIN
Query Match 88.2%; Score 680; DB 3; Length 820;
Best Local Similarity 94.6%; Pred. No. 2e-152;
Matches 704; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 21 GCCAATGTGGTTCGTGACCGTACCTCGAGTGGACACCACTTCAAGAGCCTTGAGCCAG 80
DB 1 GCCAATGTGGTTCGTGACCGTACCTCGAGTGGACACCACTTCAAGAGCCTTGAGCCAG 60
QY 81 CAGATCGAGAACATCCGGAGCCCGAGAGGAGCGGCAAGACCCCGCCGACCTCGCGT 140
DB 61 CAGATCGAGAACATCCGGAGCCCGAGAGGAGCGGCAAGACCCCGCCGACCTCGCGC 120
QY 141 GACCTCAAGATGCGCACTCTGACTGGAAGTGGAGTACTGGATTGACCCCAACCAA 200
DB 121 GACCTCAAGATGCGCACTCTGACTGGAAGTGGAGTACTGGATTGACCCCAACCAA 180
QY 201 GGCTGCACACTCGATGCCATCAAGTCTTCTGCAACATGAGACTGGTGAGACTCGGTG 260
DB 181 GGCTGCACACTCGATGCCATCAAGTCTTCTGCAACATGAGACTGGTGAGACTCGGTG 240
QY 261 TACCCCACTCAGCCAGTGTGGCCGAGAGAACTGGTACATCAGCAAGACCCCAAGGAC 320
DB 241 TACCCCACTCAGCCAGTGTGGCCGAGAGAACTGGTACATCAGCAAGACCCCAAGGAC 300
QY 321 AAGAGGCATGCTGTGGTTCGGCGGAGAGCATGACCGATGGATTCCAGTTTCGAGTATGGCGGC 380
DB 301 AAGAGGCATGCTGTGGTTCGGCGGAGAGCATGACCGATGGATTCCAGTTTCGAGTATGGCGGC 360
QY 381 CAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTCGCGCTTGATGTCCACC 440
DB 361 CAGGGCTCCGATCTGCTGACGTGGCCATCCAGCTGACCTTCTCGCGCTTGATGTCCACT 420
QY 441 GAGGCCCTCCAGAACATCATCCTACCATCTGCAAGAACAGCGTGGCTTACATGGACCGAGCAG 500
DB 421 GAGGCCCTCCAGAACATCATCCTACCATCTGCAAGAACAGCGTGGCTTACATGGACCGAGCAG 480
QY 501 ACTGGCAACTCAAGAGGCGCTGTCTCTCAAGGGCTTCAAGCGATTCAGATGATCCGCGCC 560
DB 481 ACTGGCAACTCAAGAGGCGCTGTCTCTCAAGGGCTTCAAGCGATTCAGATGATCCGCGCC 540
QY 561 GAGGGCAACACCGCTTTCACCTACAGCGTCACTGTTCGATGGCTGCAAGTCAACACCGGA 620
DB 541 GAGGGCAACACCGCTTTCACCTACAGCGTCACTGTTCGATGGCTGCAAGTCAACACCGGA 600
QY 621 GCCTGGGGCAAGACAGTGTGATGATACAAACCAAGTCTTCCCGCTGCCCATCATC 680
DB 601 GCCTGGGGCAAGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
QY 681 GATGTGGCCCCCTTGGACGTTGGTGCCCGCAGACAGGAAATTCGGCTTCGACGTTGGCCCT 740

DB 661 GATGTGGCCCCCTTGGACGTTGGCGCCCGCCGCAAGAAATTCGGCATCGACCTTAGCCCT 720
QY 741 GCTGTGCTTCTCTGTAACATCCCTCC 764
DB 721 GCTGTGCTTCTCTGTAACATCCCTCC 744
RESULT 8
AY414730 3378 bp DNA linear GSS 17-DEC-2003
LOCUS Pan troglodytes COL1A1 gene, VIRTUAL TRANSCRIPT, partial sequence,
DEFINITION genomic survey sequence.
ACCESSION AY414730.1 GI:39770692
VERSION AY414730
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Pan.
REFERENCE 1 (bases 1 to 3378)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 3378)
AUTHORS Clark,A.G., Gnanowski,S., Nielson,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
Ferriera,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source
1..3378
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>3378
/gene="COL1A1"
/locus_tag="HCM5321"
ORIGIN
Query Match 88.2%; Score 680; DB 10; Length 3378;
Best Local Similarity 92.4%; Pred. No. 2.4e-152;
Matches 683; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 17 TGAATGCCAATGTGGTTCGTGACCGTGCACCTCGAGGTGGACACCACTTCAAGAGCCTGAG 76
DB 2640 TGAATGCCAATGTGGTTCGTGACCGTGCACCTCGAGGTGGACACCACTTCAAGAGCCTGAG 2699
QY 77 CCAGCAGATCGAGAACATCCGGAGCCAGAGGAGCCGCAAGAACCCCGCCGACCTG 136
DB 2700 CCAGCAGATCGAGAACATCCGGAGCCCGAGGCGAGCCGCAAGAACCCCGCCGACCTG 2759
QY 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGATCTGGATTGACCCCAA 196
DB 2760 CCGTGACCTCAAGATGTGCCACTCTGACTGGANNNNNGAGAGTCTGGATTGACCCCAA 2819
QY 197 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 256
DB 2820 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 2879
QY 257 CGTGTACCCCACTCAGCCCCAGTGTGGCCCGAAGAACTGGTACATCAGCAAGAACCCCAA 316
DB 2880 CGTGTACCCCACTCAGCCCCAGTGTGGCCCGAAGAACTGGTACATCAGCAAGAACCCCAA 2939


```
QY 317 GGACAGAGGCGATGCTGGTTGGCGGAGAGCATGCCGATGGATTCCAGTTCGAGTATGG 376
|||||
Db 2940 GGA-NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
QY 377 CGGCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATGTC 436
|||||
Db 3000 CGGCAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCCTGCGCCTGATGTC 3059
QY 437 CACGAGGCGCTCCGAGAACATACCTACCTACCACTGCAAGACAGCGTGGCCTACATGGACCA 496
Db 3060 CACGAGGCGCTCCGAGAACATACCTACCTACCACTGCAAGACAGCGTGGCCTACATGGACCA 3119
QY 497 GCAGACTGGCACTCCAGAGGCGCTGCTCCTCAAGGCTCCACAGAGATCGAGATCCG 556
Db 3120 GCAGACTGGCACTCCAGAGGCGCTGCTCCTCAAGGCTCCACAGAGATCGAGATCCG 3179
QY 557 CGCGAGGGCAACAGCGCTTACCTACAGGTCACCTGTGATGGCTGCAGAGTCAACAC 616
Db 3180 CGCGAGGGCAACAGCGCTTACCTACAGGTCACCTGTGATGGCTGCAGAGTCAACAC 3239
QY 617 CGGAGCGCTGGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTTCCGCGCTGCCCAT 676
Db 3240 CGGAGCGCTGGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTTCCGCGCTGCCCAT 3299
QY 677 CATCGATGTGGCGCCCTTGAGCGTGTGGTGGCCAGACAGGAATTCGGCTTGGAGTGG 736
Db 3300 CATCGATGTGGCGCCCTTGAGCGTGTGGTGGCCAGACAGGAATTCGGCTTGGAGTGG 3359
QY 737 CCCTGTCTGCTTCTCTGTAA 755
Db 3360 CCCTGTCTGCTTCTCTGTAA 3378
```

RESULT 9

```
LOCUS BPI172845 874 bp mRNA linear EST 30-DEC-2003
DEFINITION BPI172845 full-length enriched swine cDNA library, adult uterus Sus
scrofa cDNA clone UTR010021G07 5', mRNA sequence.
ACCESSION BPI172845
VERSION BPI172845.1 GI:40422318
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.
1 (bases 1 to 874)
Unenishi, H., Eguchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H.,
Okumura, N., Hamasima, N. and Awata, T.
PEDE (Pig EST Data Explorer): construction of a database for ESTs
derived from porcine full-length cDNA libraries
Nucleic Acids Res. 32 (1), D484-D488 (2004)
14681463
Contact: Hirohide Unenishi
Animal Genome Laboratory, Genome Research Department
National Institute of Agrobiological Sciences
2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan
Tel: +81-29-838-8627
Fax: +81-29-838-8627
Email: huenishi@affrc.go.jp
EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute
Single pass sequencing of clones derived from oligo-capped cDNA
library
Vector sequences were eliminated by RepeatMasker version 2002/07/13
and crossmatch version 0.990319
Low quality bases were trimmed based on the quality values.
Location/Qualifiers
1..874
/organism="Sus scrofa"
/mol_type="mRNA"
```

FEATURES

source

```
RESULT 10
LOCUS CN342548 700 bp mRNA linear EST 16-MAY-2004
DEFINITION 17000424979301 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN342548
VERSION CN342548.1 GI:47342482
KEYWORDS EST.
SOURCE Homo sapiens (human)
```

```
ORIGIN
Query Match 87.5%; Score 674.8; DB 3; Length 874;
Best Local Similarity 94.3%; Pred. No. 3.5e-151;
Matches 700; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
QY 23 CAATGTGTTCTGTCGACCGTGCAGCTTCGAGGTGGACACACACCTCAAGAGCCTGAGCCAGCA 82
Db 1 CAATGTGTTCTGTCGACCGTGCAGCTTCGAGGTGGACACACACCTCAAGAGCCTGAGCCAGCA 60
QY 83 GATCGAGAAATCATCGGAGCCCGAGGGAAGCCGCAAGAACCCCGCCCGCAGCTGCGCTGA 142
Db 61 GATCGAGAAATCATCGGAGCCCGAGGAGCGAGGAGTGCAGAGAACCCCGCCCGCAGCTGCGCTGA 120
QY 143 CTTCAAGATGTGCACCTCTGACTTGGAGAGTGGAGAGTACTGGATTGACCCCAACCAAGG 202
Db 121 CTTCAAGATGTGCACCTCTGACTTGGAGAGTGGAGAGTACTGGATTGACCCCAACCAAGG 180
QY 203 CTGCAACCTGGATGCCATCAAGCTCTTCTGCAACATGGAGACTGGTGAGACCTGCCTGTA 262
Db 181 CTGCAACCTGGATGCCATCAAGCTCTTCTGCAACATGGAGACTGGTGAGACCTGCCTGTA 240
QY 263 CCCCACTCAGCCCGTGTGGCCCGAGAGAACTGGTATCATCAGCAAGAACCCCAAGGACAA 322
Db 241 CCCCACTCAGCCCGTGTGGCCCGAGAGAACTGGTATCATCAGCAAGAACCCCAAGGACAA 300
QY 323 GAGGCATGTCTGTGTTGGCGAGAGCATGACCGATGGAATTCAGTTCGAGTATGCGGCCA 382
Db 301 GAGGCATGTCTGTGTTGGCGAGAGCATGACCGAGAGCATGACCGAGATTCAGTTCGAGTATGCGGCCA 360
QY 383 GGGCTCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTGGCTGTGATGTCACCGA 442
Db 361 GGGCTCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTGGCTGTGATGTCACCGA 420
QY 443 GGCCTCCAGAACATCACCTTACCCTTGCAGAGACAGCGTGGCTTACATGACACAGCAGAC 502
Db 421 GGCCTCCAGAACATCACCTTACCCTTGCAGAGACAGCGTGGCTTACATGACACAGCAGAC 480
QY 503 TGGCAACCTCAAGAGCGCTTGTCTCTCAAGGGCTCCACGAGATCGAGATTCGCGCCGA 562
Db 481 TGGCAACCTCAAGAGCGCTTGTCTCTCAAGGGCTCCACGAGATCGAGATTCGCGCCGA 540
QY 563 GGGCAACAGCGCGTTCACCTTACAGCGTCACTGTGATGGCTGACAGTCAACCGGAGC 622
Db 541 GGGCAACAGCGCGTTCACCTTACAGCGTCACTGTGATGGCTGACAGTCAACCGGAGC 600
QY 623 CTGGGGCAAGACAGTGAATGAATACAAACACCAAGTCTCTCGCGCTGCGCCATCATCGA 682
Db 601 CTGGGGCAAGACAGTGAATGAATACAAACACCAAGTCTCTCGCGCTGCGCCATCATCGA 660
QY 683 TGTGGCGCCCTTGGAGCTTGGTTCGCCAGACCAAGGAATTCGGCTTCGACGTTGGCCCTGT 742
Db 661 TGTGGCGCCCTTGGAGCTTGGTTCGCCAGACCAAGGAATTCGGCTTCGACGTTGGCCCTGT 720
QY 743 CTGCTTCTGTAAACTCCCTCC 764
Db 721 CTGCTTCTGTAAACTCCCTCC 742
```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
1 (bases 1 to 700)
Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G. J.,
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M. S., Mandalam, R.,
Lebkowski, J. and Stanton, L. W.
Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
Nat. Biotechnol. 22 (6), 707-716 (2004)
15146197
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 700 Std Error: 0.00.

FEATURES
source Location/Qualifiers
1..700
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, embryoid bodies
derived from H1, H7 and H9 cells"
/clone_lib="GRN_EB"
/notes="oligo dt primed, full-length enriched cDNA library
from embryoid body outgrowths derived from hES cell lines
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
conditions."

ORIGIN
Query Match 87.4%; Score 674.2; DB 7; Length 700;
Best Local Similarity 99.6%; Pred. No. 4.8e-151;
Matches 676; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 17 TGATGCCAATGTGGTTCGTGACCGTGCACCTCGAGGTGGACACCACTCAAGAGCCTGAG 76
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
22 TGATGCCAATGTGGTTCGTGACCGTGCACCTCGAGGTGGACACCACTCAAGAGCCTGAG 81
QY 77 CCAGCAGATCGAAGCAATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG 136
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
82 CCAGCAGATCGAAGCAATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG 141
QY 137 CCGTGACCTCAAGATGTCACCTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 196
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
142 CCGTGACCTCAAGATGTCACCTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 201
QY 197 CCAGGCTGCAACTGGATGCCATCAAGCTCTTCTGCAACATGGAGACTGGTGGACCTG 256
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
202 CCAAGGCTGCAACTGGATGCCATCAAGCTCTTCTGCAACATGGAGACTGGTGGACCTG 261
QY 257 CGTGATACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACTAGCAAGAACCCCAA 316
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
262 CGTGATACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACTAGCAAGAACCCCAA 321
QY 317 GGAACAAGAGCATGTCTGGTTGGCGAGAGCATGACCGATGGATTCAGTTGAGTATGG 376
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
322 GGAACAAGAGCATGTCTGGTTGGCGAGAGCATGACCGATGGATTCAGTTGAGTATGG 381
QY 377 CGGCCAGGCTCGACCTCGCATGTGGCCATCCAGCTGACCTTCTGGCCTGATGTC 436
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
382 CGGCCAGGCTCGACCTCGCATGTGGCCATCCAGCTGACCTTCTGGCCTGATGTC 441
QY 437 CACGAGGCTCCAGAAACATCACTTACCACTGCAAGAACAGCGGTGGCTTACATGGACCA 496
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
442 CACGAGGCTCCAGAAACATCACTTACCACTGCAAGAACAGCGGTGGCTTACATGGACCA 501
QY 497 GCAGACTGGCAACCTCAAGAGGCTCTGCTCTTCAAGGGCTCCAAAGAGATCGAGATCGG 556
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
502 GCAGACTGGCAACCTCAAGAGGCTCTGCTCTTCAAGGGCTCCAAAGAGATCGAGATCGG 561

QY 557 CGCCGAGGGCAACAGCCGCTTACCTACAGCCTCACTGTGATGGCTGCACGAGTACAC 616
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
562 CGCCGAGGGCAACAGCCGCTTACCTACAGCCTCACTGTGATGGCTGCACGAGTACAC 621
QY 617 CGAGAGCTGGGGCAAGACAGTGTGAATACAAACACCAAGTCTCCGCCCTGCCCAT 676
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
622 CGAGAGCTGGGGCAAGACAGTGTGAATACAAACACCAAGTCTCCGCCCTGCCCAT 681
QY 677 CATCGATGTGGCCCTCTTG 695
DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
682 CATCGATGTGGCCCTCTTG 700

RESULT 11
CV547993 813 bp mRNA linear EST 03-JUN-2005
LOCUS EST409 Bovine Lambda Zap Express ovarian cortex cDNA library Bos
DEFINITION taurus cDNA clone ovca_001_b05 5', similar to Pro alpha 1 (i)
collagen, mRNA sequence.
ACCESSION CV547993 GI:66912568
VERSION CV547993.1
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Ruminantia;
Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 813)
AUTHORS Casey, O.M., Fitzpatrick, R., McInerney, J., Powell, R. and
Sreenan, J.M.
TITLE Expressed Sequence Tag analysis of abundantly expressed bovine
ovarian cortex genes
JOURNAL Unpublished (2004)
COMMENT Contact: Casey OM
Animal Reproduction Department
Teagasc, Agriculture and Food Development Authority
Galway, Ireland
Tel: 00353 91 845845
Fax: 00353 91 845847
Email: ocasey@athenry.teagasc.ie
Insert Length: 813 Std Error: 0.00
Plate: ovarian cortex abundant plate 1 row: b column: 05
Seq primer: M13 reverse primer = caggaacagctatgacc
High quality sequence stop: 813
POLYA=No. Location/Qualifiers
1..813
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="ovca_001_b05"
/sex="Female"
/tissue_type="Ovarian cortex"
/dev_stage="Day 2, 6, 8 and 14 of the oestrus cycle"
/clone_lib="Bovine Lambda Zap Express ovarian cortex cDNA
library"
/note="Organ: Ovarian cortex; Vector: Lambda Zap Express;
Site 1: EcoRI; Site 2: XhoI; A bovine ovarian cortex cDNA
library was constructed using the Lambda Zap
Express/gigapack cloning kit (Stratagene cloning systems).
cDNA synthesis was carried out using an oligo-(dT) primer
for the reverse transcription of 5(g of mRNA and the
library was constructed by directional cloning EcoRI-XhoI
based on manufacturers instructions. An insert:vector
ligation ratio of 1:5 was chosen as most optimum. The
lambda library was packaged with Gigapack III gold
packaging extracts and plated on the E. coli cell line
XL1-Blue MRF'."

ORIGIN
Query Match 87.4%; Score 673.6; DB 7; Length 813;
Best Local Similarity 95.3%; Pred. No. 6.8e-151;

Matches 705; Conservative 0; Mismatches 34; Indels 1; Gaps 1;
Qy 29 GGTTCGTGACCGTGAACCTCGAGGTGGACACACCTCTCAAGAGCTGTGACCGAGCAGATCGA 88
Db 2 GGTTCGTGACCGTGAACCTCGAGGTGGACACACCTCTCAAGAGCTGTGACCGAGCAGATCGA 61
Qy 89 GAAATCCGGAGCCGAGAGGAGCGGCAAGAACCCCGCCGACCTGCGGTGACCTCAA 148
Db 62 GAAATCCGGAGCCGAGAGGAGCGGCAAGAACCCCGCCGACCTGCGGTGACCTCAA 121
Qy 149 GATGTGCCATCTCACTGGAGAGTGGAGATCTGGAATGACCCCAACCAAGCTGCAA 208
Db 122 GATGTGCCATCTCACTGGAGAGCGGAGATCTGGAATGACCCCAACCAAGCTGCAA 181
Qy 209 CTTGGATGCCATCAAGTCTTCTGCAATGAGACCTGTGTGAGACCTGCGTGTACCCAC 268
Db 182 CTTGGATGCCATCAAGTCTTCTGCAATGAGAACCGGTGAGACCTGTGTATACCCAC 241
Qy 269 TCAGCCAGTGTGCCGAGAGAACTGTAATCATCAGCAAGAACCCCAAGGACAGAGGCA 328
Db 242 TCAGCCAGTGTGCCGAGAGAACTGTAATCATCAGCAAGAACCCCAAGGAGAGGCA 301
Qy 329 TGTCTGTGTTGGCGAGAGCATGACCGATGATTCAGTTCAGTATGCGCGCCAGGGCTC 388
Db 302 CGTCTGTGTTGGCGAGAGCATGACCGCGGATTCAGTTCAGTATGCGCGCCAGGGCTC 361
Qy 389 CGACCTTGGCGAGTGGCCATCCAGCTGACCTTCTGCGCTGATGTCCACCGAGGCTC 448
Db 362 CGATCTTGGCGAGTGGCCATCCAGCTGACCTTCTGCGCTGATGTCCACCGAGGCTC 421
Qy 449 CCAGAACATCACTACCACTGCAGAACAGCGTGGCTTACATGACAGCAGATGAGCA 508
Db 422 CCAGAACATCACTACCACTGCAGAACAGCGTGGCTTACATGACAGCAGATGAGCA 481
Qy 509 CCTCAAGAGGCCCTGTCTCTCAAGGGCTCCAAAGAGATCGAGATCCGCGCCAGGGCAA 568
Db 482 CCTCAAGAGGCCCTGTCTCTCAAGGGCTCCAAAGAGATCGAGATCCGCGCCAGGGCAA 541
Qy 569 CAGCGCTTCACTACAGCGTCACTGTGATGCTGACAGTCCAGTCCACCGAGCCTGGG 628
Db 542 CAGCGCTTCACTACAGCGTCACTGTGATGCTGACAGTCCAGTCCACCGAGCCTGGG 601
Qy 629 CAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCTGCGCATCATGATGTCG 688
Db 602 CAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCTGCGCATCATGATGTCG 661
Qy 689 CCGCTTGGAGCTTGGTGGCCCGAGACCAAGGAATTCGCTTTCGAGTGGCGCTGTCTGCT 747
Db 662 CCGCTTGGAGCTTGGTGGCCCGAGACCAAGGAATTCGCTTTCGAGTGGCGCTGTCTGCT 721
Qy 748 TCCTGTAACTCCCTCCATC 767
Db 722 TCCTGTAACTCCCTCCATC 741

RESULT 12

BP172993

LOCUS

DEFINITION BP172993 full-length enriched swine cDNA library, adult uterus Sus

scrofa cDNA clone UTR010038G10 5', mRNA sequence.

ACCESSION BP172993

VERSION BP172993.1

KEYWORDS EST.

SOURCE Sus scrofa (pig)

ORGANISM

Sus scrofa (Bases 1 to 800)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;
Sus.

REFERENCE

AUTHORS

Uniehi, H., Enuchi, T., Suzuki, K., Sawazaki, T., Toki, D., Shinkai, H.,
Okumura, N., Hamasina, N. and Awata, T.
PSE (Pig EST Data Explorer): construction of a database for ESTs
derived from porcine full-length cDNA libraries

JOURNAL

PUBMED

COMMENT

Nucleic Acids Res. 32 (1), D484-D488 (2004)

14681463

Contact: Hirohide Uniehi

Animal Genome Laboratory, Genome Research Department

National Institute of Agrobiological Sciences

2 Ikenodai, Tsukuba, Ibaraki 305-8602, Japan

Tel: +81-29-838-8627

Fax: +81-29-838-8627

Email: huenishi@affrc.go.jp

EST project with full-length enriched cDNA libraries carried out in
Animal Genome Research Program (Japan) by National Institute of
Agrobiological Sciences and STAFF-Institute

Single pass sequencing of clones derived from oligo-capped cDNA

library sequences were eliminated by RepeatMasker version 2002/07/13

and crossmatch version 0.990319

Low quality bases were trimmed based on the quality values.

FEATURES

Location/Qualifiers

source

1..800

/organism="Sus scrofa"

/mol_type="mRNA"

/db_xref="taxon:9823"

/clone="UTR010038G10"

/tissue_type="uterus"

/dev_stage="adult"

/clone_lib="full-length enriched swine cDNA library, adult

uterus"

ORIGIN

Query Match 87.3%; Score 673; DB 3; Length 800;

Best Local Similarity 94.6%; Pred. No. 9.4e-151;

Matches 697; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 28 TGGTTCGTGACCGTGAACCTCGAGGTGGACACCACTCTCAAGAGCTGTGACGATCG 87

Db 1 TGGTTCGTGACCGTGAACCTCGAGGTGGACACCACTCTCAAGAGCTGTGACGATCG 60

Qy 88 AGAACATCCGAGAGCGGAGGAGCGGCAAGAACCCCGCCGACCTGCGGTGACCTCA 147

Db 61 AGAACATCCGAGAGCGGAGGAGCGGCAAGAACCCCGCCGACCTGCGGTGACCTCA 120

Qy 148 AGATGTGCCACTCTGATGGAAGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCA 207

Db 121 AGATGTGCCACTCTGATGGAAGAGTGGAGAGTACTGGATTGACCCCAACCAAGGCTGCA 180

Qy 208 ACCTGTGATGTCATCAAGTCTTCTGCAACATGAGAGTGTGTGAGACCTGCGGTGACCCCA 267

Db 181 ACCTGTGATGTCATCAAGTCTTCTGCAACATGAGAGTGTGTGAGACCTGCGGTGACCCCA 240

Qy 268 CTGAGCCGAGTGTGGCCCGAGAGAACTGGTATCATCAGCAAGAACCCCAAGGAGGC 327

Db 241 CTGAGCCGAGTGTGGCCCGAGAGAACTGGTATCATCAGCAAGAACCCCAAGGAGGC 300

Qy 328 ATGTCTGTTGCGGAGAGCATGACCGATGATTCAGTTCGAGTATGCGGCGGAGGCT 387

Db 301 ACCTGTGTTGCGGAGAGCATGACCGATGATTCAGTTCGAGTATGCGGCGGAGGCT 360

Qy 388 CCGACCTGCGGATGTGGCCATCCAGTGTGACCTTCTGCGCTGTGATCCACCGAGGCT 447

Db 361 CCGATCTGCTGAGTGGCCATCCAGTGTGACCTTCTGCGCTGTGATCCACCGAGGCT 420

Qy 448 CCAGAACATCACTACCTACCTGCAAGAACAGCGTGGCTTACATGGACCAAGGAGGC 507

Db 421 CCAGAACATCACTACCTACCTGCAAGAACAGCGTGGCTTACATGGACCAAGGAGGC 480

Qy 508 ACCTCAAGAGGCGCTGCTCTTCAAGGCTTCAAGAGATCGAGATCCGCGGAGGCA 567

Db 481 ACCTCAAGAGGCGCTGCTCTTCAAGGCTTCAAGAGATCGAGATCCGCGGAGGCA 540

Qy 568 ACAGCGCTTCACTACAGGCTCACTGTGATGGCTGCAAGTGTGACCGAGCTGGG 627

Db 541 ACAGCGCTTCACTACAGGCTCACTGTGATGGCTGCAAGTGTGACCGAGCTGGG 600

```

QY 628 GCAAGACAGTGATGTAATCAAAACACACCAAGTCCTCCCGCCTGCCCATCATCGATGTGG 687
Db 601 GCAAGACAGTGATGTAATCAAAACACACCAAGTCCTCCCGCCTGCCCATCATCGATGTGG 660
QY 688 CCCCCTTGGACGTTGGTGGCCGAGACCAAGAAATTCCGCTTCGACGTTGGCCCTGTCTGCT 747
Db 661 CCCCCTTGGACGTTGGTGGCCGAGACCAAGAAATTCCGCTTCGACGTTGGCCCTGTCTGCT 720
QY 748 TCCTGTAAACTCCCTCC 764
Db 721 TCCTGTAAACTCCCTCC 737

RESULT 13
CD607437/c
LOCUS 796 bp mRNA linear EST 12-JAN-2004
DEFINITION 5609595371 FLP Homo sapiens cDNA, mRNA sequence.
ACCESSION CD607437
VERSION CD607437.1 GI:40255700
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 796)
AUTHORS Fu, G.K., Wang, J.T., Yang, J., Au-Young, J. and Stuve, L.L.
TITLE Circular rapid amplification of cDNA ends for high-throughput
extension cloning of partial genes
JOURNAL Genomics 84 (1), 205-210 (2004)
PUBMED 15203218
COMMENT Contact: Fu GK
Incyte Genomics, Inc.
3160 Porter Dr., Palo Alto, CA 94304, USA
Tel: 6508454102
Email: gfu@incyte.com.
FEATURES
source
1..796
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="FLP"
/notes="Vector: pDrive Cloning Vector"
ORIGIN
Query Match 87.2%; Score 672; DB 6; Length 796;
Best Local Similarity 99.0%; Pred. No. 1.6e-150;
Matches 697; Conservative 0; Mismatches 5; Indels 2; Gaps 2;
QY 17 TGATGCCAATGTGGTTGTCGTCACCGTGA-CCTCGAGGTGGAC-ACCACCTCAAGAGCCTG 74
Db 718 TGATGCCAATGTGGTTGTCGTCACCGTGA-CCTCGAGGTGGAC-ACCACCTCAAGAGCCTG 659
QY 75 AGCCAGCAGATCGAGAACATCCGAGCCCGCAGAGGGAAGCCGCAAGAACCCCGCCGACCC 134
Db 658 AGCCAGCAGATCGAGAACATCCGAGCCCGCAGAGGCGAGCGCAGAACCCCGCCGACCC 599
QY 135 TGGCGTGACCTCAAGATGTCGCACTGCTGACTGAAGAGTGAGAGTACTGATTGACCCC 194
Db 598 TGGCGTGACCTCAAGATGTCGCACTGCTGACTGAAGAGTGAGAGTACTGATTGACCCC 539
QY 195 AACCAAGGCTGCACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACC 254
Db 538 AACCAAGGCTGCACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACC 479
QY 255 TGGGTGTACCCCACTCAGCCCACTGTGGCCCGCAGAGAACTGGTACATCAGCAAGAACCCC 314
Db 478 TGGGTGTACCCCACTCAGCCCACTGTGGCCCGCAGAGAACTGGTACATCAGCAAGAACCCC 419
QY 315 AAGCAACAGAGGCAATGTCGTGGTTCGGCGAGAGATGACCGATGGATTCGAGTTCGAGTAT 374
Db 418 AAGCAACAGAGGCAATGTCGTGGTTCGGCGAGAGATGACCGATGGATTCGAGTTCGAGTAT 359

```

```

QY 375 GGGGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTTGGCCCTGATG 434
Db 358 GGGGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTTGGCCCTGATG 299
QY 435 TTCACCGAGGCTCCCGAGAACATCACTACCACTGCAAGAACAGCGTGGCCCTACATGGAC 494
Db 298 TCACCGAGGCTCCCGAGAACATCACTACCACTGCAAGAACAGCGTGGCCCTACATGGAC 239
QY 495 CAGCAGACTGGCAACCTCAAGAGGCGCTTGCCTCAAGGGCTCCAAACGAGATCGAGATC 554
Db 238 CAGCAGACTGGCAACCTCAAGAGGCGCTTGCCTCAAGGGCTCCAAACGAGATCGAGATC 179
QY 555 CGCGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATGGCTGCGAGAGTCAC 614
Db 178 CGCGCCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGATGGCTGCGAGAGTCAC 119
QY 615 ACCGAGGCTGGGGCAAGACAGTGAATTGAATACAAACACCAAGTCCTCCCGCCCTGCC 674
Db 118 ACCGAGGCTGGGGCAAGACAGTGAATTGAATACAAACACCAAGTCCTCCCGCCCTGCC 59
QY 675 ATCATCGATGTGGCCCGCTTGGACGTTGGTGGCCCGCAGACGGA 718
Db 58 ATCATCGATGTGGCCCGCTTGGACGTTGGTGGCCCGCAGACGGA 15

RESULT 14
CN342240
LOCUS 701 bp mRNA linear EST 16-MAY-2004
DEFINITION 17000532189205 GRN_EB Homo sapiens cDNA 5', mRNA sequence.
ACCESSION CN342240
VERSION CN342240.1 GI:47342174
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominidae; Homo.
REFERENCE 1 (bases 1 to 701)
AUTHORS Brandenberger, R., Wei, H., Zhang, S., Lei, S., Murage, J., Fisk, G.J.,
Li, Y., Xu, C., Fang, R., Guegler, K., Rao, M.S., Mandalam, R.,
Lebowski, J. and Stanton, L.W.
TITLE Transcriptome characterization elucidates signaling networks that
control human ES cell growth and differentiation
JOURNAL Nat. Biotechnol. 22 (6), 707-716 (2004)
PUBMED 15146197
COMMENT Contact: Brandenberger R
Regenerative Medicine
Geron Corporation
230 Constitution Drive, Menlo Park, CA 94025, USA
Tel: 650 473 8658
Fax: 650 473 7760
Email: rbrandenberger@geron.com
Insert Length: 701 Std Error: 0.00.
FEATURES
source
1..701
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="embryonic stem cells, embryoid bodies
derived from H1, H7 and H9 cells"
/clone_lib="GRN EB"
/notes="Oligo dT primed, full-length enriched cDNA library
from embryoid body outgrowths derived from HES cell lines
H1 (p32), H7 (p29), and H9 (p26) maintained in feeder-free
conditions."
ORIGIN
Query Match 87.1%; Score 671.2; DB 7; Length 701;
Best Local Similarity 99.6%; Pred. No. 2.5e-150;
Matches 673; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 17 TGATGCCAATGTGGTTGTCGTCACCGTGA-CCTCGAGGTGGAC-ACCACCTCAAGAGCCTGAG 76

```

Db 26 TGATGCCAATGTGGTTCGTGACCGTGTGACCTCGAGTGGACACCCCTCAAGACGCTGAG 85
Qy 77 CCAGCAGATCCAGAACATCCGGAGCCCGAGAGGAGCGCGAAGACCCCGCCCGCAGACTG 136
Db 86 CCAGCAGATCCAGAACATCCGGAGCCCGAGAGGAGCGCGAAGACCCCGCCCGCAGACTG 145
Qy 137 CCGTGACCTCAAGATGSCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 146 CCGTGACCTCAAGATGSCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 205
Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTTCGCAACATGGAGACTGGTGAGACCTG 256
Db 206 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTTCGCAACATGGAGACTGGTGAGACCTG 265
Qy 257 CGTGATACCCCACTCAGCCCACTGTGGCCGAGAGAACTGGTACATCAAGAAACCCCAA 316
Db 266 CGTGATACCCCACTCAGCCCACTGTGGCCGAGAGAACTGGTACATCAAGAAACCCCAA 325
Qy 317 GGACAAGAGGCATGTCTGGTTGGCGAGAGCATGACCGATGATTCAGATTCCAGATTGG 376
Db 326 GGACAAGAGGCATGTCTGGTTGGCGAGAGCATGACCGATGATTCAGATTGG 385
Qy 377 CGGCACAGGCTCCGACCTCGGATGTGGCCATCCAGCTGACCTTCTTGGCCCTGATGTC 436
Db 386 CGGCACAGGCTCCGACCTCGGATGTGGCCATCCAGCTGACCTTCTTGGCCCTGATGTC 445
Qy 437 CACGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCGTGCGCTACATGGACCA 496
Db 446 CACGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCGTGCGCTACATGGACCA 505
Qy 497 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGCTCCAGAGATCGAGATCCG 556
Db 506 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGCTCCAGAGATCGAGATCCG 565
Qy 557 CGCGAGGCGCAACAGCGCTTCACTCAGAGGTCACTGTGATGGCTGCAAGATCACAC 616
Db 566 CGCGAGGCGCAACAGCGCTTCACTCAGAGGTCACTGTGATGGCTGCAAGATCACAC 625
Qy 617 CGGAGCTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCCCGCTGCCCAT 676
Db 626 CGGAGCTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCCCGCTGCCCAT 685
Qy 677 CATCATGTGGCCCCC 692
Db 686 CATCATGTGGCCCCC 701

RESULT 15

DR432524/c

LOCUS

DEFINITION

AGENCOURT 51294715 DFCI-Vidal horfeome 1.1 Homo sapiens cDNA clone

IMAGE:40029743 3', mRNA sequence.

DR432524

DR432524.1 GI:68341061

EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;

Hominidae; Homo.

1 (bases 1 to 696)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

NTH-MGC <http://mgc.nci.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Daniela S. Gerhard, Ph.D.

Office of Cancer Genomics

National Cancer Institute / NIH

Bldg. 31 Rm10A07 Bethesda, MD 20892

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Marc Vidal

cDNA Library Preparation: Dana Farber Cancer Institute/Vidal Lab

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LML)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LML at: <http://image.llnl.gov>

Plate: IRBY1 row: g column: 08

High quality sequence stop: 628.

Location/Qualifiers

source

1..696

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:40029743"

/lab_host="DH10B Tora"

/clone_lib="DFCI-Vidal horfeome 1.1"

/notes="Organ: mixed; Vector: pDONR223; Site_1: attB1.1;

Site_2: attB2.1; Genome Rep. 2004 Oct;14(10B):2128-35."

ORIGIN

Query Match 87.0%; Score 671; DB 8; Length 696;
Best Local Similarity 98.4%; Pred. No. 2.8e-150;
Matches 677; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Qy 67 AGAGCCTGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCG 126
Db 693 AGAGCCTGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCG 634
Qy 127 CCAGCAGCTCCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGA 186
Db 633 CCAGCAGCTCCGCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGA 574
Qy 187 TTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGAGACTG 246
Db 573 TTGACCCCAACCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGAGACTG 514
Qy 247 GTGAGACTCTCGTGTATCCCACTCAGCCAGTGTGGGCCAGAGAACTGGTACATCAGCA 306
Db 513 GTGAGACTCTCGTGTATCCCACTCAGCCAGTGTGGGCCAGAGAACTGGTACATCAGCA 454
Qy 307 AGAACCCCAAGGCAAGAGGCAATGTCTGGTTCGGCGAGAGCATGACCGATGGATTCAGT 366
Db 453 AGAACCCCAAGGCAAGAGGCAATGTCTGGTTCGGCGAGAGCATGACCGATGGATTCAGT 394
Qy 367 TCAGATATGGCGGCCAGGCTCCGACCTCCGAGTGTGGCCATCCAGTCACTGACCTTCTTCG 426
Db 393 TCAGATATGGCGGCCAGGCTCCGACCTCCGAGTGTGGCCATCCAGTCACTGACCTTCTTCG 334
Qy 427 GCCTGATGTCCACCGAGGCTCCGAGAACTCACTACCTACCAAGAACAGCAGCTGGCCT 486
Db 333 GCCTGATGTCCACCGAGGCTCCGAGAACTCACTACCTACCAAGAACAGCAGCTGGCCT 274
Qy 487 ACATGACACGACAGCTGGCAACTCAAGAGGCTTCCTCAAGGGCTTCAAGGCTTCAACGAGA 546
Db 273 ACATGACACGACAGCTGGCAACTCAAGAGGCTTCCTCAAGGGCTTCAAGGCTTCAACGAGA 214
Qy 547 TCAGATATGGCGGCCAGGCTCCGACCTCCGAGTGTGGCCATCCAGTCACTGATGGCTGCA 606
Db 213 TCAGATATGGCGGCCAGGCTCCGACCTCCGAGTGTGGCCATCCAGTCACTGATGGCTGCA 154
Qy 607 CGAGTCAACCGAGCTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCTCCC 666
Db 153 CGAGTCAACCGAGCTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCTCCC 94
Qy 667 GCCTGCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCAAGACAGAGAAATCGGCT 726
Db 93 GCCTGCCCATCATCGATGTGGCCCCCTTGGACGTTGGTGGCCCAAGACAGAGAAATCGGCT 34
Qy 727 TCAGAGTGGCCCTGTCTGCTTCTGTA 754
Db 33 TCAGAGTGGCCATGTCTGCTTCTGTA 6

Search completed: April 27, 2006, 22:40:06

Job time : 3928.36 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 18:34:23 ; Search time 161.848 Seconds
(without alignments)
8467.835 Million cell updates/sec

Title: US-10-677-877A-3

Perfect score: 771

Sequence: 1 aagctacgaagatctgat.....gtaactccctccatctaga 771

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/H COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCPUS COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/PP COMB.seq:*
- 8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
- 9: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	738.8	95.8	4409	3	US-09-331-347C-22
2	390.4	50.6	590	3	US-09-404-879A-215
3	390.4	50.6	590	3	US-09-338-933-215
4	390.4	50.6	590	3	US-09-215-681-215
5	390.4	50.6	590	3	US-09-216-003A-215
6	390.4	50.6	590	3	US-09-667-857-215
7	390.4	50.6	590	3	US-10-198-053-215
8	389.2	50.5	594	3	US-09-827-271-215
9	389.2	50.5	594	3	US-09-404-879A-214
10	389.2	50.5	594	3	US-09-338-933-214
11	389.2	50.5	594	3	US-09-215-681-214
12	389.2	50.5	594	3	US-09-216-003A-214
13	389.2	50.5	594	3	US-09-667-857-214
14	389.2	50.5	594	3	US-10-198-053-214
15	389.2	50.5	594	3	US-09-827-271-214
16	378.6	49.1	704	3	US-08-896-164-49
17	372.4	48.3	544	3	US-09-404-879A-257
18	372.4	48.3	544	3	US-09-338-933-257
19	372.4	48.3	544	3	US-09-215-681-257
20	372.4	48.3	544	3	US-09-216-003A-257
21	372.4	48.3	544	3	US-09-667-857-257
22	372.4	48.3	544	3	US-10-198-053-257
23	372.4	48.3	544	3	US-09-827-271-257
24	367.2	47.6	587	3	US-09-404-879A-192

ALIGNMENTS

RESULT 1

US-09-331-347C-22
; Sequence 22, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Me
; TITLE OF INVENTION: obtaining Such and Their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331,347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 22
; LENGTH: 4409
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-331-347C-22

Query Match	95.8%	Score 738.8;	DB 3;	Length 4409;
Best Local Similarity	99.7%	Pred. No. 1.4e-165;		
Matches	740;	Conservative	0;	Mismatches 2;
Indels	0;	Gaps	0;	
Qy	17	TGATGCCAATGTGGTTTCGTGACCGTGCAGCTCGAGGTGGACCAACCCCTCAAGAGCCTGAG	76	Sequence 192, App
Db	3667	TGATGCCAATGTGGTTTCGTGACCGTGCAGCTCGAGGTGGACCAACCCCTCAAGAGCCTGAG	3736	Sequence 192, App
Qy	77	CGAGCAGATCGAGAAATCCGAGGCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG	136	Sequence 192, App
Db	3727	CGAGCAGATCGAGAAATCCGAGGCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG	3786	Sequence 258, App
Qy	137	CCGTGACCTCAAGATGTCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA	196	Sequence 258, App
Db	3787	CCGTGACCTCAAGATGTCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA	3846	Sequence 258, App
Qy	197	CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG	256	Sequence 258, App
Db	3847	CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG	3906	Sequence 258, App
Qy	257	CGTGTACCCCACTCAGCCCACTGGTGGCCCAAGAACTGTATCATCAGCAAGAACCCCAA	316	Sequence 257, App
Db	3907	CGTGTACCCCACTCAGCCCACTGGTGGCCCAAGAACTGTATCATCAGCAAGAACCCCAA	3966	Sequence 257, App
Qy	317	GGACAAAGGCGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	376	Sequence 418, App
Db	3967	GGACAAAGGCGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	4026	Sequence 418, App

377 CGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTGTGGCCTGATGTC 436
4027 CGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTGTGGCCTGATGTC 4086
437 CACCGAGGCTCCCGAAGCAATACCTACCACTGCAAGAAACAGCGTGGCCTTACATGGACCA 496
4087 CACCGAGGCTCCCGAAGCAATACCTACCACTGCAAGAAACAGCGTGGCCTTACATGGACCA 4146
497 GCAGACTGGCAACTCTAAGAGGCTTCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556
4147 GCAGACTGGCAACTCTAAGAGGCTTCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 4206
557 CGCGAGGGCAACAGCGCTTACCTACAGGTCACGTGCTCGATGGTGCACAGATCACAC 616
4207 CGCGAGGGCAACAGCGCTTACCTACAGGTCACGTGCTCGATGGTGCACAGATCACAC 4266
617 CGGAGCTCGGGCAAGACAGTGTGAATGAATAAACAACCAAGTCTCTCCGCGCTGCCCAT 676
4267 CGGAGCTCGGGCAAGACAGTGTGAATGAATGAATAAACAACCAAGTCTCTCCGCGCTGCCCAT 4326
677 CATCGATGGGCCCCCTTGAAGCTTGGTGGCCCGAGACAGGAATTCGGCTTCGACGTTGG 736
4327 CATCGATGGGCCCCCTTGAAGCTTGGTGGCCCGAGACAGGAATTCGGCTTCGACGTTGG 4386
737 CCCTGTCTGCTTCTCTGTAAGT 758
4387 CCCTGTCTGCTTCTCTGTAAGT 4408

RESULT 2
US-09-404-879A-215
; Sequence 215, Application US/09404879A
; Patent No. 648546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/404,879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(590)
; OTHER INFORMATION: n = A,T,C or G
US-09-404-879A-215

Query Match 50.6%; Score 390.4; DB 3; Length 590;
Best Local Similarity 98.5%; Pred. No. 3.3e-83;
Matches 394; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

17 TGATGCCAATGTGGTTCGTGACCGTGTGACCTCGAGGTGGACACCAACCTTCAAGAGCCTGAG 76
180 TGATGCCAATGTGGTTCGTGACCGTGTGACCTCGAGGTGGACACCAACCTTCAAGAGCCTGAG 239
77 CCAGCAGATCGAGAACATCCGGAGCCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG 136
240 CCAGCAGATCGAGAACATCCGGAGCCCGAGGCGCAGCGCAAGAACCCCGCCGACCTG 299
137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTCTGGATTGACCCCAA 196
300 CCGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTCTGGATTGACCCCAA 359
197 CCAAGGCTGCAACCTCGATGCCATCAAGTCTTCTGCAACATCGAGACTGGTGAGACCTG 256
360 CCAAGGCTGCAACCTCGATGCCATCAAGTCTTCTGCAACATCGAGACTGGTGAGACCTG 419
257 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAACTGGTATCATCAGCAAGAACCCCAA 316
420 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAACTGGTATCATCAGCAAGAACCCCAA 479
317 GGCACAGAGGCATGTCTGGTTCGGCCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 136
480 GGCACAGAGGCATGTCTGGTTCGGCCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
480 GGCACAGAGGCATGTCTGGTTCGGCCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 539
377 CGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTG 416
540 CGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCGCGCG 579

Db 360 CCAAGGCTGCAACCTCGATGCCATCAAGTCTTCTGCAACATCGAGACTGGTGAGACCTG 419
QY 257 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAACTGGTATCATCAGCAAGAACCCCAA 316
Db 420 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAACTGGTATCATCAGCAAGAACCCCAA 479
QY 317 GGCACAGAGGCATGTCTGGTTCGGCCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
Db 480 GGCACAGAGGCATGTCTGGTTCGGCCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 539
QY 377 CGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTG 416
Db 540 CGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCGCGCG 579

RESULT 3
US-09-338-933-215
; Sequence 215, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; TITLE OF INVENTION: OVARIAN CANCER
; FILE REFERENCE: 210121.462C1
; CURRENT APPLICATION NUMBER: US/09/338,933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(590)
; OTHER INFORMATION: n = A,T,C or G
US-09-338-933-215

Query Match 50.6%; Score 390.4; DB 3; Length 590;
Best Local Similarity 98.5%; Pred. No. 3.3e-83;
Matches 394; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

17 TGATGCCAATGTGGTTCGTGACCGTGTGACCTCGAGGTGGACACCAACCTTCAAGAGCCTGAG 76
180 TGATGCCAATGTGGTTCGTGACCGTGTGACCTCGAGGTGGACACCAACCTTCAAGAGCCTGAG 239
77 CCAGCAGATCGAGAACATCCGGAGCCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG 136
240 CCAGCAGATCGAGAACATCCGGAGCCCGAGGCGCAGCGCAAGAACCCCGCCGACCTG 299
137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTCTGGATTGACCCCAA 196
300 CCGTGACCTCAAGATGTGCCACTCTGACTGGAGAGTGGAGAGTCTGGATTGACCCCAA 359
197 CCAAGGCTGCAACCTCGATGCCATCAAGTCTTCTGCAACATCGAGACTGGTGAGACCTG 256
360 CCAAGGCTGCAACCTCGATGCCATCAAGTCTTCTGCAACATCGAGACTGGTGAGACCTG 419
257 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAACTGGTATCATCAGCAAGAACCCCAA 316
420 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAACTGGTATCATCAGCAAGAACCCCAA 479
317 GGCACAGAGGCATGTCTGGTTCGGCCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
480 GGCACAGAGGCATGTCTGGTTCGGCCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 539
377 CGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTG 416
540 CGGCCAGGGCTCCGACCCCTGCGATGTGGCCATCCGCGCG 579

RESULT 4

```
US-09-215-681-215
; Sequence 215, Application US/09215681A
; Patent No. 6528253
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF OVARIAN CANCER
; FILE REFERENCE: 210121.463
; CURRENT APPLICATION NUMBER: US/09/215.681A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(590)
; OTHER INFORMATION: n = A,T,C or G
US-09-215-681-215

Query Match          50.6%; Score 390.4; DB 3; Length 590;
Best Local Similarity 98.5%; Pred. No. 3.3e-83;
Matches 394; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 17 TGATGCCAATGTGTTCTGTCGACCGTGCAGCTCGAGGTGGACACACCTCAAGAGCCTGAG 76
Db 180 TGATGCCAATGTGTTCTGTCGACCGTGCAGCTCGAGGTGGACACACCTCAAGAGCCTGAG 239

Qy 77 CCAGCAGATCGAGAAACATCCGGAGCCCGAGAGGAGCGGCAAGAACCCGCGCCGACCTG 136
Db 240 CCAGCAGATCGAGAAACATCCGGAGCCCGAGAGGAGCGGCAAGAACCCGCGCCGACCTG 299

Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 300 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 359

Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 256
Db 360 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 419

Qy 257 CGTGATACCCCACTCAGCCCACTGAGTGGCCCAAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 420 CGTGATACCCCACTCAGCCCACTGAGTGGCCCAAGAGAACTGGTACATCAGCAAGAACCCCAA 479

Qy 317 GGACAGAGGCGATGCTGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
Db 480 GGACAGAGGCGATGCTGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 539

Qy 377 CGGCCAGGGCTCCGACCCCTGCCGATGTGGCCATCCAGCTG 416
Db 540 CGGCCAGGGCTCCGACCCCTGCCGATGTGGACCTCCGGCCG 579

RESULT 6
US-09-667-857-215
; Sequence 215, Application US/09667857
; Patent No. 6695664
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C5
; CURRENT APPLICATION NUMBER: US/09/667,857
; CURRENT FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 215
; LENGTH: 590
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(590)
; OTHER INFORMATION: n = A,T,C or G

US-09-216-003A-215
; Sequence 215, Application US/09216003A
; Patent No. 6670463
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF OVARIAN CANCER
; TITLE OF INVENTION: OF OVARIAN CANCER
; FILE REFERENCE: 210121.462
; CURRENT APPLICATION NUMBER: US/09/216.003A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 215
; LENGTH: 590
```


Qy 377 CGGCCAGGGCTCCGACCTCGCGATGGCCATCCAGCTG 416
|||||
Db 540 CGGCCAGGGCTCCGACCTCGCGATGGCCATCCAGCTG 579
|||||

RESULT 9

US-09-404-879A-214/c
; Sequence 214, Application US/09404879A
; Patent No. 6468546
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C2
; CURRENT APPLICATION NUMBER: US/09/404.879A
; CURRENT FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 393
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 214
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(594)
; OTHER INFORMATION: n = A,T,C or G
US-09-404-879A-214

Query Match 50.5%; Score 389.2; DB 3; Length 594;
Best Local Similarity 99.2%; Pred. No. 6.3e-83;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 17 TGATGCCAATGTGTTTCGTGACCTCGAGTGGACACACCTCAAGAGCCTGAG 76
|||||
Db 409 TGATGCCAATGTGTTTCGTGACCTCGAGTGGACACACCTCAAGAGCCTGAG 350
|||||
Qy 77 CCAGCAGATCGAAGAACATCCGAGCCCGAGAGGAGCGGCAAGAACCCCGCCGACCTG 136
|||||
Db 349 CCAGCAGATCGAAGAACATCCGAGCCCGAGAGGAGCGGCAAGAACCCCGCCGACCTG 290
|||||
Qy 137 CCCTGACCTCAAGATGTCACCTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 196
|||||
Db 289 CCCTGACCTCAAGATGTCACCTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 230
|||||
Qy 197 CCAAGGCTGCAACTGCGATGCCATCAAGTCTTCTGCAACATGGAGACTGTGAGACCTG 256
|||||
Db 229 CCAAGGCTGCAACTGCGATGCCATCAAGTCTTCTGCAACATGGAGACTGTGAGACCTG 170
|||||
Qy 257 CGTGTACCCCACTCAGCCCACTGTCGGCCCGAGAGCATGCGATGGATTCCAGTTCCAGTATGG 316
|||||
Db 169 CGTGTACCCCACTCAGCCCACTGTCGGCCCGAGAGCATGCGATGGATTCCAGTTCCAGTATGG 110
|||||
Qy 317 GGACAAGAGGATGTCGTGTTCCGCGAGAGCATGCGATGGATTCCAGTTCCAGTATGG 376
|||||
Db 109 GGACAAGAGGATGTCGTGTTCCGCGAGAGCATGCGATGGATTCCAGTTCCAGTATGG 50
|||||
Qy 377 CGGCCAGGGCTCCGACCTCGCGATGGCCATC 410
|||||
Db 49 CGGCCAGGGCTCCGACCTCGCGATGGACCTC 16
|||||

RESULT 10

US-09-338-933-214/c
; Sequence 214, Application US/09338933
; Patent No. 6488931
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer Lynn
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF
; TITLE OF INVENTION: OVARIAN CANCER
; FILE REFERENCE: 210121.462C1

; CURRENT APPLICATION NUMBER: US/09/338,933
; CURRENT FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 312
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 214
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(594)
; OTHER INFORMATION: n = A,T,C or G
US-09-338-933-214

Query Match 50.5%; Score 389.2; DB 3; Length 594;

Best Local Similarity 99.2%; Pred. No. 6.3e-83;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGTTTCGTGACCTCGAGTGGACACACCTCAAGAGCCTGAG 76
|||||
Db 409 TGATGCCAATGTGTTTCGTGACCTCGAGTGGACACACCTCAAGAGCCTGAG 350
|||||
Qy 77 CCAGCAGATCGAAGAACATCCGAGCCCGAGAGGAGCGGCAAGAACCCCGCCGACCTG 136
|||||
Db 349 CCAGCAGATCGAAGAACATCCGAGCCCGAGAGGAGCGGCAAGAACCCCGCCGACCTG 290
|||||
Qy 137 CCCTGACCTCAAGATGTCACCTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 196
|||||
Db 289 CCCTGACCTCAAGATGTCACCTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 230
|||||
Qy 197 CCAAGGCTGCAACTGCGATGCCATCAAGTCTTCTGCAACATGGAGACTGTGAGACCTG 256
|||||
Db 229 CCAAGGCTGCAACTGCGATGCCATCAAGTCTTCTGCAACATGGAGACTGTGAGACCTG 170
|||||
Qy 257 CGTGTACCCCACTCAGCCCACTGTCGGCCCGAGAGCATGCGATGGATTCCAGTTCCAGTATGG 316
|||||
Db 169 CGTGTACCCCACTCAGCCCACTGTCGGCCCGAGAGCATGCGATGGATTCCAGTTCCAGTATGG 110
|||||
Qy 317 GGACAAGAGGATGTCGTGTTCCGCGAGAGCATGCGATGGATTCCAGTTCCAGTATGG 376
|||||
Db 109 GGACAAGAGGATGTCGTGTTCCGCGAGAGCATGCGATGGATTCCAGTTCCAGTATGG 50
|||||
Qy 377 CGGCCAGGGCTCCGACCTCGCGATGGCCATC 410
|||||
Db 49 CGGCCAGGGCTCCGACCTCGCGATGGACCTC 16
|||||

RESULT 11

US-09-215-681-214/c
; Sequence 214, Application US/09215681A
; Patent No. 6528253
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Frudakis, Tony N.
; APPLICANT: King, Gordon E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSIS
; TITLE OF INVENTION: OF OVARIAN CANCER
; FILE REFERENCE: 210121.463
; CURRENT APPLICATION NUMBER: US/09/215,681A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 214
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(594)
; OTHER INFORMATION: n = A,T,C or G
US-09-215-681-214

Query Match 50.5%; Score 389.2; DB 3; Length 594;

```
Best Local Similarity 99.2%; Pred. No. 6.3e-83;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 17 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACACCTCAAGAGCCTGAG 76
Db 409 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACACCTCAAGAGCCTGAG 350
QY 77 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTTG 136
Db 349 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTTG 290
QY 137 CCGTGACCTCAAGATGCGCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 289 CCGTGACCTCAAGATGCGCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 230
QY 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 170
Db 229 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 170
QY 257 CGTGATACCCACTCAGCCAGTGTGCCAGAGACTGGTACATCAGCAAGAACCCCAA 316
Db 169 CGTGATACCCACTCAGCCAGTGTGCCAGAGACTGGTACATCAGCAAGAACCCCAA 110
QY 317 GGACAAGAGGCATGTCTGGTTCGGCGAGAGCATGCCAGTGGATTCCAGTTCGAGTATGG 376
Db 109 GGACAAGAGGCATGTCTGGTTCGGCGAGAGCATGCCAGTGGATTCCAGTTCGAGTATGG 50
QY 377 CGGCCAGGGCTCGAGCCCTCGATGGGCCATC 410
Db 49 CGGCCAGGGCTCGAGCCCTCGATGGGCCATC 16

RESULT 12
US-09-216-003A-214/c
; Sequence 214, Application US/09216003A
; Patent No. 6670463
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: King, Gordon E.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steven P.
; APPLICANT: Retter, Steven W.
; APPLICANT: Fanger, Gary Richard
; APPLICANT: Reed, Steven G.
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF OVARIAN CANCER
; FILE REFERENCE: 210121.462
; CURRENT APPLICATION NUMBER: US/09/216.003A
; CURRENT FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 310
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 214
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (452)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (509)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (585)
; OTHER INFORMATION: Where n is a, c, g or t
US-09-216-003A-214

Query Match 50.5%; Score 389.2; DB 3; Length 594;
Best Local Similarity 99.2%; Pred. No. 6.3e-83;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 17 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACCTCAAGAGCCTGAG 76
Db 409 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACCTCAAGAGCCTGAG 350
QY 77 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTTG 136
Db 349 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTTG 290
QY 137 CCGTGACCTCAAGATGCGCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 289 CCGTGACCTCAAGATGCGCACTCTGACTGGAGAGTGGAGAGTACTGGATTGACCCCAA 230
QY 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 256
Db 229 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 170

Query Match 50.5%; Score 389.2; DB 3; Length 594;
Best Local Similarity 99.2%; Pred. No. 6.3e-83;
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 17 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACCTCAAGAGCCTGAG 76
Db 409 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACACCTCAAGAGCCTGAG 350
QY 77 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTTG 136
Db 349 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTTG 290
```

Qy	257	CGTGTAACCCCACTCAGCCCAAGTGTGGCCAGAGAACTGGTACATCAGCAAGAAACCCCAA	316
Db	169	CGTGTAACCCCACTCAGCCCAAGTGTGGCCAGAGAACTGGTACATCAGCAAGAAACCCCAA	110
Qy	317	GGACAAAGAGGCATGCTGGTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	376
Db	109	GGACAAAGAGGCATGCTGGTTCGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG	50
Qy	377	CGGCCAGAGGCTCCGACCCCTGCCGATGTGCCCATC	410
Db	49	CGGCCAGAGGCTCCGACCCCTGCCGATGTGGACCTC	16

RESULT 14

US-10-198-053-214/c

Sequence 214, Application US/10198053

Patent No. 6858710

GENERAL INFORMATION:

APPLICANT: Bangur, Chaitanya S.

APPLICANT: Better, Marc W.

APPLICANT: Fanger, Gary R.

APPLICANT: Hill, Paul

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

FILE REFERENCE: 210121.462C9

CURRENT APPLICATION NUMBER: US/10/198,053

CURRENT FILING DATE: 2002-07-17

NUMBER OF SEQ ID NOS: 624

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 214

LENGTH: 594

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: 452, 509, 585

OTHER INFORMATION: n = A,T,C or G

US-10-198-053-214

```

US-09-827-271-214/c
; Sequence 214, Application US/09827271
; Patent No. 6963980
; GENERAL INFORMATION:
; APPLICANT: Retter, Marc W.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.462C6
; CURRENT APPLICATION NUMBER: US/09/827,271
; CURRENT FILING DATE: 2001-04-04
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 214
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(594)
; OTHER INFORMATION: n = A, T, C or G
US-09-827-271-214

Query Match          50.5%; Score 389.2; DB 3; Length 594;
Best Local Similarity 99.2%; Pred. No. 6.3e-83;
Matches 391; Conservative 0; Mismatches 3; Indels 0;

Qy 17 TGATGCCAATGTGGTTGGTGCACCGTGACCTCGAGGTGGACACCAACCCCTCAAGAG
Db 409 TGATGCCAATGTGGTTGGTGCACCGTGACCTCGAGGTGGACACCAACCCCTCAAGAG
Qy 77 CCAGCAGATCGAGAACATCCGGAGCCCGAGAGGGAAGCCGCAAGAACCCCGCCGCC
Db 349 CCAGCAGATCGAGAACATCCGGAGCCCGAGAGGCGAGCGCAGAACCCCGCCGCC
Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTCGAAGAGTGGAGAGTACTGGATTGGA
Db 289 CCGTGACCTCAAGATGTGCCACTCTGACTCGAAGAGTGGAGAGTACTGGATTGGA
Qy 197 CCAAGGTCGAACTCGATGCCATCAAGTCTTCTGCAACATGAGAGACTGGTGA
Db 229 CCAAGGTCGAACTCGATGCCATCAAGTCTTCTGCAACATGAGAGACTGGTGA
Qy 257 CGTGTACCCCACTCAGCCCAAGTGTGGCCGAGAAGAACTGGTACATCAGCAAGAA
Db 169 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAGAACTGGTACATCAGCAAGAA
Qy 317 GGAACAAGAGGCACTCTGTGTTTCGGCCGAGAGCATGACCCGATGGATTCAGTTTCGA
Db 109 GGAACAAGAGGCACTCTGTGTTTCGGCCGAGAGCATGACCCGATGGATTCAGTTTCGA
Qy 377 CGGCCGAGGGCTCGAACCTTCGCCGATGTGGCCATC 410
Db 49 CGGCCGAGGGCTCGAACCTTCGCCGATGTGGACCTC 16

```

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 18:53:19 ; Search time 697.19 Seconds
(without alignments)
9144.835 Million cell updates/sec

Title: US-10-677-877A-3

Perfect score: 771

Sequence: 1 aagcttcagtaagatctgat.....gtaaacctccatcatctaga 771

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	771	100.0	771	9	US-10-677-877A-3
2	760	98.6	1947	9	US-10-677-877A-15
3	760	98.6	2294	9	US-10-677-877A-7
4	759	98.4	1542	9	US-10-677-877A-11
5	755	97.9	963	9	US-10-677-877A-1
6	755	97.9	1734	9	US-10-677-877A-9
7	755	97.9	2139	9	US-10-677-877A-13
8	755	97.9	2487	9	US-10-677-877A-5
9	751	97.4	5921	6	US-10-177-293-66
10	751	97.4	5921	6	US-10-301-822-29
11	751	97.4	5921	7	US-10-784-425-18
12	751	97.4	5921	7	US-10-734-564-12
13	751	97.4	5921	8	US-10-788-792-23
14	751	97.4	5921	8	US-10-723-860-2288
15	751	97.4	5921	9	US-10-956-157-340
16	751	97.4	5921	9	US-10-287-436A-139
17	751	97.4	6728	3	US-09-954-456-782
18	751	97.4	6728	3	US-09-880-107-3946
19	751	97.4	6728	3	US-09-918-715-260
20	751	97.4	6728	3	US-09-873-367C-521
21	751	97.4	6728	3	US-09-873-367C-1066
22	751	97.4	6728	5	US-10-060-036-136
23	751	97.4	6728	5	US-10-171-311-35

24	751	97.4	6728	5	US-10-149-352-1	Sequence 1, Appli
25	751	97.4	6728	6	US-10-177-293-64	Sequence 64, Appli
26	751	97.4	6728	6	US-10-301-822-27	Sequence 27, Appli
27	751	97.4	6728	8	US-10-474-794-260	Sequence 260, App
28	751	97.4	6728	9	US-10-843-641A-521	Sequence 521, App
29	751	97.4	6728	9	US-10-843-641A-1066	Sequence 1066, Ap
30	751	97.4	6728	9	US-10-843-641A-3809	Sequence 3809, Ap
31	751	97.4	6728	9	US-10-852-335A-63	Sequence 63, Appli
32	751	97.4	6728	9	US-10-979-159-260	Sequence 260, App
33	751	97.4	6728	9	US-10-287-436A-755	Sequence 755, App
34	746.2	96.8	4744	6	US-10-084-817-49	Sequence 49, Appli
35	746.2	96.8	4770	6	US-10-291-265-7	Sequence 7, Appli
36	746.2	96.8	4791	6	US-10-291-265-479	Sequence 479, App
37	746.2	96.8	6739	5	US-10-116-802-237	Sequence 237, App
38	746.2	96.8	6753	5	US-10-198-846-13252	Sequence 13252, A
39	745.8	96.7	2192	3	US-09-925-301-42	Sequence 42, Appli
40	745.8	96.7	3347	5	US-10-000-773A-16	Sequence 16, Appli
41	739	95.8	778	6	US-10-139-794-80	Sequence 80, Appli
42	738.8	95.8	4409	5	US-10-216-705-22	Sequence 22, Appli
43	735.2	95.4	6769	5	US-10-198-846-13897	Sequence 13897, A
44	725.6	94.1	1746	5	US-10-198-846-13901	Sequence 13901, A
45	723	93.8	1684	8	US-10-723-860-4837	Sequence 4837, Ap

ALIGNMENTS

RESULT 1
US-10-677-877A-3
; Sequence 3, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng / GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 3
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(755)
US-10-677-877A-3

Query Match		100.0%;	Score 771;	DB 9;	Length 771;
Best Local Similarity		100.0%;	Pred. No. 9.1e-215;		
Matches 771;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	AAGCTTACGTAAGATCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGTGGACACCA	60		
Db	1	AAGCTTACGTAAGATCTGATGCCAATGTGGTTCGTGACCGTGACCTCGAGTGGACACCA	60		
Qy	61	CCCTCAAGAGCCTTGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGA	120		
Db	61	CCCTCAAGAGCCTTGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGA	120		
Qy	121	ACCCGCGCGACCTCGCGTGACTCAGATGTCACCTGACTGGAAGAGTGGAGAGT	180		
Db	121	ACCCGCGCGACCTCGCGTGACTCAGATGTCACCTGACTGGAAGAGTGGAGAGT	180		
Qy	181	ACTGGATTGACCCCAACCAAGGCTGCAACCTGATGCCATCAAAAGTCTTCTGCAACATGG	240		
Db	181	ACTGGATTGACCCCAACCAAGGCTGCAACCTGATGCCATCAAAAGTCTTCTGCAACATGG	240		
Qy	241	AGACTGGTGAAGCTCGGTGTATCCCACTCAGCCCAAGTGTGGCCCAAGAACTGGTACA	300		
Db	241	AGACTGGTGAAGCTCGGTGTATCCCACTCAGCCCAAGTGTGGCCCAAGAACTGGTACA	300		
Qy	301	TCAGCAAGAAACCCCAAGGACAGAGGATGTCTGTGTTTCGCGGAGAGCATGACCGATGGAT	360		

```
Db 301 TCAGCAAGAACCCCAAGGCAATGCTGTTGGCGGAGAGCATGACCGATGGAT 360
Qy 361 TCAGATTTCAGATGCGCGGCGGCTCCGACCTCGCGGATGCGCATCCAGCTGACCT 420
Db 361 TCAGATTTCAGATGCGCGGCGGCTCCGACCTCGCGGATGCGCATCCAGCTGACCT 420
Qy 421 TCCTGCGCTGATGTCACCGAGGCTCCGAGAACATCACTTACCACTGCAAGAACAGCG 480
Db 421 TCCTGCGCTGATGTCACCGAGGCTCCGAGAACATCACTTACCACTGCAAGAACAGCG 480
Qy 481 TGCGCTACATGGAACGAGCAGATGCGCAACTCAAGAGGCGCTGCTCCTCAAGGGCTCCA 540
Db 481 TGCGCTACATGGAACGAGCAGATGCGCAACTCAAGAGGCGCTGCTCCTCAAGGGCTCCA 540
Qy 541 ACAGATCGAGATCGCGCGGCGGCAAGCGCTTCACTACAGGTCACCTGTCGATG 600
Db 541 ACAGATCGAGATCGCGCGGCGGCAAGCGCTTCACTACAGGTCACCTGTCGATG 600
Qy 601 GCTGCAAGATTCACACCGAGGCTTGGGGCAAGACAGTGAATTGAATACAAAACCAACAGT 660
Db 601 GCTGCAAGATTCACACCGAGGCTTGGGGCAAGACAGTGAATTGAATACAAAACCAACAGT 660
Qy 661 CCTCCGCTGCGCCATCATGATGTCGCGCTTGGACGTTGGTGGCCCAAGACAGGAAT 720
Db 661 CCTCCGCTGCGCCATCATGATGTCGCGCTTGGACGTTGGTGGCCCAAGACAGGAAT 720
Qy 721 TCGGCTTCGAGCTTGGCGCTGCTGCTTCTGTAACTCCCTCCATCTAGA 771
Db 721 TCGGCTTCGAGCTTGGCGCTGCTGCTTCTGTAACTCCCTCCATCTAGA 771
```

RESULT 2

```
US-10-677-877A-15
; Sequence 15, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 15
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (24)...(1931)
US-10-677-877A-15
```

```
Query Match 98.6%; Score 760; DB 9; Length 1947;
Best Local Similarity 100.0%; Pred. No. 1.8e-211;
Matches 760; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 12 AGATCTGATCCCATGTTGGTTCTGACCGTGAACCTCGAGGTGACACCACTTCAAGAGC 71
Db 1188 AGATCTGATCCCATGTTGGTTCTGACCGTGAACCTCGAGGTGACACCACTTCAAGAGC 1247
Qy 72 CTGAGCCAGCAGATCGAGAACATCCGAGGCCGAGGGAAGCGCAAGAACCCGCCCGC 131
Db 1248 CTGAGCCAGCAGATCGAGAACATCCGAGGCCGAGGGAAGCGCAAGAACCCGCCCGC 1307
Qy 132 ACCTGCGGTGACCTCAAGATGTGCACTCTGACTGGAAGAGTGGAGATGATGATGAC 191
Db 1308 ACCTGCGGTGACCTCAAGATGTGCACTCTGACTGGAAGAGTGGAGATGATGATGAC 1367
Qy 192 CCNACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTTCGCAACATGGAGACTGGTGAG 251
Db 1368 CCNACCAAGGCTGCAACCTGGATGCCATCAAAGTCTTTCGCAACATGGAGACTGGTGAG 1427
```

```
Qy 252 ACCTGCGGTGATCCCACTCAGCCCACTGAGTGGGCCCAAGAACTGGTACATCAGCAAGAAC 311
Db 1428 ACCTGCGGTGATCCCACTCAGCCCACTGAGTGGGCCCAAGAACTGGTACATCAGCAAGAAC 1487
Qy 312 CCNACCAAGCAGAGGCACTGCTGTTTCGGGCGAGAGCATGACGATGGATTCAGATTTCGAG 371
Db 1488 CCNACCAAGCAGAGGCACTGCTGTTTCGGGCGAGAGCATGACGATGGATTCAGATTTCGAG 1547
Qy 372 TATGGCGGCGCAGGGGCTCCGACCTCGCGGATGTGGCCCATCAGCTGACCTTCTTGCSCCTG 431
Db 1548 TATGGCGGCGCAGGGGCTCCGACCTCGCGGATGTGGCCCATCAGCTGACCTTCTTGCSCCTG 1607
Qy 432 ATGTCCACCGAGGCTTCCAGAACATCACTTACCACTGCAAGAAACAGCGTGGCTTACATG 491
Db 1608 ATGTCCACCGAGGCTTCCAGAACATCACTTACCACTGCAAGAAACAGCGTGGCTTACATG 1667
Qy 492 GACCACGACACTGGCAACCTCAAGAGGCGCTTCTCCTCAAGGGCTCCCAACGAGATCGAG 551
Db 1668 GACCACGACACTGGCAACCTCAAGAGGCGCTTCTCCTCAAGGGCTCCCAACGAGATCGAG 1727
Qy 552 ATCCGCGCGCAGGGCAACAGCGCTTCACTTACAGCGTCACTGTGATGGCTGCACGAGT 611
Db 1728 ATCCGCGCGCAGGGCAACAGCGCTTCACTTACAGCGTCACTGTGATGGCTGCACGAGT 1787
Qy 612 CACACCGAGGCTTGGGGCAAGACAGTGAATTGAATACAAAACCAAGTCTTCCCGCCTG 671
Db 1788 CACACCGAGGCTTGGGGCAAGACAGTGAATTGAATACAAAACCAAGTCTTCCCGCCTG 1847
Qy 672 CCNATCATGATGTCGCGCTTGGAGCTTGGTGGCCCGAGACAGGAATTCGGCTTCGAC 731
Db 1848 CCNATCATGATGTCGCGCTTGGAGCTTGGTGGCCCGAGACAGGAATTCGGCTTCGAC 1907
Qy 732 GTTGGCGCTGCTGCTTCTGTAACTCCCTCCATCTAGA 771
Db 1908 GTTGGCGCTGCTGCTTCTGTAACTCCCTCCATCTAGA 1947
```

RESULT 3

```
US-10-677-877A-7
; Sequence 7, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 7
; LENGTH: 2294
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2278)
US-10-677-877A-7
```

```
Query Match 98.6%; Score 760; DB 9; Length 2294;
Best Local Similarity 100.0%; Pred. No. 1.9e-211;
Matches 760; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 12 AGATCTGATGCCAATGTTGGTTCTGACCGTGAACCTCGAGGTGACACCACTTCAAGAGC 71
Db 1535 AGATCTGATGCCAATGTTGGTTCTGACCGTGAACCTCGAGGTGACACCACTTCAAGAGC 1594
Qy 72 CTGAGCCAGCAGATCGAGAACATCCGAGGCCGAGGGAAGCGCAAGAACCCGCCCGC 131
Db 1595 CTGAGCCAGCAGATCGAGAACATCCGAGGCCGAGGGAAGCGCAAGAACCCGCCCGC 1654
Qy 132 ACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGACTGGATGAC 191
Db 1655 ACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGACTGGATGAC 1714
```

QY 192 CCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGAGACTGGTGAG 251
DB 1715 CCCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGAGACTGGTGAG 1774
QY 252 ACCTGCGTGTACCCACCTACGACCCAGTGTGCGCCAGAGAACTGTGATCATCAGCAAGAAC 311
DB 1775 ACCTGCGTGTACCCACCTACGACCCAGTGTGCGCCAGAGAACTGTGATCATCAGCAAGAAC 1834
QY 312 CCCAAGGACAAAGAGGCTATGTCTGGTTCGGCGAGAGCATGACCCGATGATTCAGTTCGAG 371
DB 1835 CCCAAGGACAAAGAGGCTATGTCTGGTTCGGCGAGAGCATGACCCGATGATTCAGTTCGAG 1894
QY 372 TATGCGGCGCAGGCGCTCGACCTTCCGATGTGCGCATTCAGCTGACCTTCTCGGCGCTG 431
DB 1895 TATGCGGCGCAGGCGCTCGACCTTCCGATGTGCGCATTCAGCTGACCTTCTCGGCGCTG 1954
QY 432 ATGTCCACCGAGGCTTCCGAGAACATCACCTACCACTGCAAGACAGCGTGGCCTACATG 491
DB 1955 ATGTCCACCGAGGCTTCCGAGAACATCACCTACCACTGCAAGACAGCGTGGCCTACATG 2014
QY 492 GACCAGCAGACTGCAACCTCAAGAGGCGCTGCTCTCAAGGCTCCAAACGAGATCGAG 551
DB 2015 GACCAGCAGACTGCAACCTCAAGAGGCGCTGCTCTCAAGGCTCCAAACGAGATCGAG 2074
QY 552 ATCCGCGCGAGGCGCAACAGCCGCTTCACTACAGCGTCACTGTCGATGCGTGCACGAGT 611
DB 2075 ATCCGCGCGAGGCGCAACAGCCGCTTCACTACAGCGTCACTGTCGATGCGTGCACGAGT 2134
QY 612 CACACCGGAGCGTGGGCGAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCGTG 671
DB 2135 CACACCGGAGCGTGGGCGAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCGTG 2194
QY 672 CCATCATCGATGTGGCGCCCTTGGAGTGGTGGCCCGACAGCAAGAAATTCGGCTTCGAC 731
DB 2195 CCATCATCGATGTGGCGCCCTTGGAGTGGTGGCCCGACAGCAAGAAATTCGGCTTCGAC 2254
QY 732 GTTGGCCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 771
DB 2255 GTTGGCCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 2294

RESULT 4

US-10-677-877A-11
; Sequence 11, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 11
; LENGTH: 1542
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18)...(1526)
US-10-677-877A-11

Query Match 98.4%; Score 759; DB 9; Length 1542;
Best Local Similarity 100.0%; Pred. No. 3.4e-211;
Matches 759; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 GATCTGATGCAATGTGGTTCGTGACCGTCACTCGAGTGGACACCACTCAAGAGCC 72
DB 784 GATCTGATGCAATGTGGTTCGTGACCGTCACTCGAGTGGACACCACTCAAGAGCC 843
QY 73 TGAGCCAGCAGATCGAGAACATCCGGAGCCCGAGAGGAAACCGCAAGAACCCCGCCGCA 132

DB 844 TGAGCCAGCAGATCGAGAACATCCGGAGCCCGAGAGGAGCGCAAGAACCCCGCCCGCA 903
QY 133 CCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAAGAGTGGAGACTACTGGATTGACC 192
DB 904 CCTGCCGTGACCTCAAGATGTGCCACTCTGACTGGAAAGAGTGGAGACTACTGGATTGACC 963
QY 193 CCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGAGACTGGTGAGA 252
DB 964 CCAACCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGAGACTGGTGAGA 1023
QY 253 CTTGCGTGTACCCACCTACGACCCAGTGTGCGCCAGAGAACTGTGATCATCAGCAAGAAC 312
DB 1024 CTTGCGTGTACCCACCTACGACCCAGTGTGCGCCAGAGAACTGTGATCATCAGCAAGAAC 1083
QY 313 CCAAGGACAAAGAGGCTATGTCTGGTTCGGCGAGAGCATGACCCGATGATTCAGTTCGAGT 372
DB 1084 CCAAGGACAAAGAGGCTATGTCTGGTTCGGCGAGAGCATGACCCGATGATTCAGTTCGAGT 1143
QY 373 ATGCGCGCGCAGGCGCTCGACCTTCCGATGTGCGCATTCAGCTGACCTTCTCGGCGCTGA 432
DB 1144 ATGCGCGCGCAGGCGCTCGACCTTCCGATGTGCGCATTCAGCTGACCTTCTCGGCGCTGA 1203
QY 433 TGTCCACCGAGGCGCTCCAGAACATCACCTACCACTGCAAGAAACAGCGTGGCCTACATGG 492
DB 1204 TGTCCACCGAGGCGCTCCAGAACATCACCTACCACTGCAAGAAACAGCGTGGCCTACATGG 1263
QY 493 ACCAGCAGACTGCGCAACCTCAAGAGGCGCTGCTCTCAAGGCTCCAAACGAGATCGAGA 552
DB 1264 ACCAGCAGACTGCGCAACCTCAAGAGGCGCTGCTCTCAAGGCTCCAAACGAGATCGAGA 1323
QY 553 TCCGCGCGAGGCGCAACAGCCGCTTCACTACAGCGTCACTGTCGATGGCTGCAACGAGTC 612
DB 1324 TCCGCGCGAGGCGCAACAGCCGCTTCACTACAGCGTCACTGTCGATGGCTGCAACGAGTC 1383
QY 613 ACACCGGAGCGCTGGGCGAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCGCTGC 672
DB 1384 ACACCGGAGCGCTGGGCGAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCGCTGC 1443
QY 673 CCATCATCGATGTGGCGCCCTTGGAGTGGTGGCCCGACAGCAAGAAATTCGGCTTCGACG 732
DB 1444 CCATCATCGATGTGGCGCCCTTGGAGTGGTGGCCCGACAGCAAGAAATTCGGCTTCGACG 1503
QY 733 TTGCGCCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 771
DB 1504 TTGCGCCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 1542

RESULT 5

US-10-677-877A-1
; Sequence 1, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 1
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(947)
US-10-677-877A-1

Query Match 97.9%; Score 755; DB 9; Length 963;
Best Local Similarity 100.0%; Pred. No. 4.5e-210;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 TGATGCCAATGTGGTTCGTGACCGTTCGAGTGGACCACTCGAGGTGGACCACTCCCAAGAGCCTGAG 76

```
Db 209 TGATGCCAATGTGGTTCTGACCGTGCACCTCGAGGTGGACACACCCCTCAAGAGCCCTGAG 268
QY 77 CCAGCAGATCGAGAAATATCGGAGCCAGAGGGAAGCCGCAAGAAACCCCGCCGACCTG 136
Db 269 CCAGCAGATCGAGAAATATCGGAGCCAGAGGGAAGCCGCAAGAAACCCCGCCGACCTG 328
QY 137 CCGTGACCTCAAGATGTGCCATCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 329 CCGTGACCTCAAGATGTGCCATCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 388
QY 197 CCAAGGCTGCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 256
Db 389 CCAAGGCTGCAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 448
QY 257 CCGTGATGCCCACTCAGCCAGTGTGCCCAAGAGAACTGGTATCATCAGCAAGAAACCCCAA 316
Db 449 CCGTGATGCCCACTCAGCCAGTGTGCCCAAGAGAACTGGTATCATCAGCAAGAAACCCCAA 508
QY 317 GGAACAAGAGGATGTCTGGTTTCGGCGAGAGCATGACCGATGATTCAGTTCGAGTATGG 376
Db 509 GGAACAAGAGGATGTCTGGTTTCGGCGAGAGCATGACCGATGATTCAGTTCGAGTATGG 568
QY 377 CGGCCAAGGCTCCGACCTCGCATGTGCGCATTCAGCTGACCTTCTCTGGCGCTGATGTC 436
Db 569 CGGCCAAGGCTCCGACCTCGCATGTGCGCATTCAGCTGACCTTCTCTGGCGCTGATGTC 628
QY 437 CACCGAGGCTCCGAGAAATACCTACCTACCTCAAGAGGCTCCAAAGAGTCCAGATCCG 496
Db 629 CACCGAGGCTCCGAGAAATACCTACCTACCTCAAGAGGCTCCAAAGAGTCCAGATCCG 688
QY 497 GCAGACTGGCAACTCAAGAGGCTTCCTCTCAAGAGGCTCCAAAGAGTCCAGATCCG 556
Db 689 GCAGACTGGCAACTCAAGAGGCTTCCTCTCAAGAGGCTCCAAAGAGTCCAGATCCG 748
QY 557 CGCCGAGGCAACAGCGCTTCACTACAGCTGCTGCTGATGATGATGATGATGATGATGATG 616
Db 749 CGCCGAGGCAACAGCGCTTCACTACAGCTGCTGCTGATGATGATGATGATGATGATGATG 808
QY 617 CGAGGCTGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCGCTGCCCAT 676
Db 809 CGAGGCTGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCGCTGCCCAT 868
QY 677 CATCGATGTGGCCCTTGGACCTTGGTGGCCAGAGTCCAGGAAATTCGGCTTCGACGTTGG 736
Db 869 CATCGATGTGGCCCTTGGACCTTGGTGGCCAGAGTCCAGGAAATTCGGCTTCGACGTTGG 928
QY 737 CCTGTCTGCTTCTCTGTAACCTCCCTCCATCTAGA 771
Db 929 CCTGTCTGCTTCTCTGTAACCTCCCTCCATCTAGA 963
```

```
RESULT 6
US-10-677-877A-9
; Sequence 9, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 9
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (18) ... (1718)
US-10-677-877A-9
```

```
Query Match 97.9%; Score 755; DB 9; Length 1734;
Best Local Similarity 100.0%; Pred. No. 5.1e-210; Indels 0; Gaps 0;
Matches 755; Conservative 0; Mismatches 0;

QY 17 TGATGCCAATGTGGTTCTGACCGTGCACCTCGAGGTGGACACACCCCTCAAGAGCCTGAG 76
Db 980 TGATGCCAATGTGGTTCTGACCGTGCACCTCGAGGTGGACACACCCCTCAAGAGCCTGAG 1039
QY 77 CCAGCAGATCGAGAAATATCGGAGCCAGAGGGAAGCCGCAAGAAACCCCGCCGACCTG 136
Db 1040 CCAGCAGATCGAGAAATATCGGAGCCAGAGGGAAGCCGCAAGAAACCCCGCCGACCTG 1099
QY 137 CCGTGACCTCAAGATGTGCCATCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 1100 CCGTGACCTCAAGATGTGCCATCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 1159
QY 197 CCAAGGCTCCAACTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
Db 1160 CCAAGGCTCCAACTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 1219
QY 257 CCGTGATGCCCACTCAGCCAGTGTGGCCCAAGAGAACTGGTATCATCAGCAAGAAACCCCAA 316
Db 1220 CCGTGATGCCCACTCAGCCAGTGTGGCCCAAGAGAACTGGTATCATCAGCAAGAAACCCCAA 1279
QY 317 GGAACAAGAGGATGTCTGGTTTCGGCGAGAGCATGACCGATGATTCAGTTCGAGTATGG 376
Db 1280 GGAACAAGAGGATGTCTGGTTTCGGCGAGAGCATGACCGATGATTCAGTTCGAGTATGG 1339
QY 377 CGGCCAAGGCTCCGAGAAATACCTACCTACCTCAAGAGGCTCCAAAGAGTCCAGATCCG 436
Db 1340 CGGCCAAGGCTCCGAGAAATACCTACCTACCTCAAGAGGCTCCAAAGAGTCCAGATCCG 1399
QY 437 CACCGAGGCTCCGAGAAATACCTACCTACCTCAAGAGAAACAGCGTGGCTTACATGGACCA 496
Db 1400 CACCGAGGCTCCGAGAAATACCTACCTACCTCAAGAGAAACAGCGTGGCTTACATGGACCA 1459
QY 497 GCAGACTGGCAACTCAAGAGGCTTGTCTCAAGGCTCCAGAGAGTCCAGAGATCGAGATCCG 556
Db 1460 GCAGACTGGCAACTCAAGAGGCTTGTCTCAAGGCTCCAGAGAGTCCAGAGATCGAGATCCG 1519
QY 557 CGCCGAGGCAACAGCGCTTCACTACAGCTGCTGCTGATGATGATGATGATGATGATGATG 616
Db 1520 CGCCGAGGCAACAGCGCTTCACTACAGCTGCTGCTGATGATGATGATGATGATGATGATG 1579
QY 617 CGAGGCTGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCGCTGCCCAT 676
Db 1580 CGAGGCTGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCGCTGCCCAT 1639
QY 677 CATCGATGTGGCCCTTGGACGTTGGTGGCCAGAGTCCAGGAAATTCGGCTTCGACGTTGG 736
Db 1640 CATCGATGTGGCCCTTGGACGTTGGTGGCCAGAGTCCAGGAAATTCGGCTTCGACGTTGG 1699
QY 737 CCTGTCTGCTTCTCTGTAACCTCCCTCCATCTAGA 771
Db 1700 CCTGTCTGCTTCTCTGTAACCTCCCTCCATCTAGA 1734
```

RESULT 7

```
US-10-677-877A-13
; Sequence 13, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 13
; LENGTH: 2139
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (24)...(2123)
US-10-677-877A-13

Query Match      97.9%; Score 755; DB 9; Length 2139;
Best Local Similarity 100.0%; Pred. No. 5.4e-210;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 TGATGCCAATGTGGTTCGTGACCGTGTGACCGTGTGAGTGGACACCAACCCCTCAAGAGCCTGAG 76
Db 1385 TGATGCCAATGTGGTTCGTGACCGTGTGACCGTGTGAGTGGACACCAACCCCTCAAGAGCCTGAG 1444

Qy 77 CCAGCAGATCGAGAACATCCGAGCCCGAGAGGAGCGCGAAGAACCCCGCCGCACTG 136
Db 1445 CCAGCAGATCGAGAACATCCGAGCCCGAGAGGAGCGCGAAGAACCCCGCCGCACTG 1504

Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA 196
Db 1505 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA 1564

Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 256
Db 1565 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 1624

Qy 257 CGTGATACCCCACTCAGCCCAAGTGTGGCCCGAGAGAACTGGTACATCAGAGAAACCCCAA 316
Db 1625 CGTGATACCCCACTCAGCCCAAGTGTGGCCCGAGAGAACTGGTACATCAGAGAAACCCCAA 1684

Qy 317 GGAACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGATTCGAGTATGG 376
Db 1685 GGAACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGATTCGAGTATGG 1744

Qy 377 CGGCGAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTCGCGCCCTGATGTC 436
Db 1745 CGGCGAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTCGCGCCCTGATGTC 1804

Qy 437 CACCGAGGCTCCGAGAACATCACTACCACTGCAAGAAACAGCGTGGCCCTACATGGACCA 496
Db 1805 CACCGAGGCTCCGAGAACATCACTACCACTGCAAGAAACAGCGTGGCCCTACATGGACCA 1864

Qy 497 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556
Db 1865 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 1924

Qy 557 CGCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGAGTCAAC 616
Db 1925 CGCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGAGTCAAC 1984

Qy 617 CGGAGCTGGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCCCTGCCAT 676
Db 1985 CGGAGCTGGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCCCTGCCAT 2044

Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGGCCCGAGACAGGAATTCGGCTTCGACGTTGG 736
Db 2045 CATCGATGTGGCCCTTGGACGTTGGTGGCCCGAGACAGGAATTCGGCTTCGACGTTGG 2104

Qy 737 CCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 771
Db 2105 CCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 2139
```

```

RESULT 8
US-10-677-877A-5
; Sequence 5, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; TITLE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677.877A
; CURRENT FILING DATE: 2003-10-02
```

```

; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 5
; LENGTH: 2487
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12)...(2471)
US-10-677-877A-5

Query Match      97.9%; Score 755; DB 9; Length 2487;
Best Local Similarity 100.0%; Pred. No. 5.5e-210;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 TGATGCCAATGTGGTTCGTGACCGTGTGACCGTGTGAGTGGACACCAACCCCTCAAGAGCCTGAG 76
Db 1733 TGATGCCAATGTGGTTCGTGACCGTGTGACCGTGTGAGTGGACACCAACCCCTCAAGAGCCTGAG 1792

Qy 77 CCAGCAGATCGAGAACATCCGAGCCCGAGAGGAGCGCGAAGAACCCCGCCGCACTG 136
Db 1793 CCAGCAGATCGAGAACATCCGAGCCCGAGAGGAGCGCGAAGAACCCCGCCGCACTG 1852

Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA 196
Db 1853 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTTGGATTGACCCCAA 1912

Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 256
Db 1913 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACCTG 1972

Qy 257 CGTGATACCCCACTCAGCCCAAGTGTGGCCCGAGAGAACTGGTACATCAGAGAAACCCCAA 316
Db 1973 CGTGATACCCCACTCAGCCCAAGTGTGGCCCGAGAGAACTGGTACATCAGAGAAACCCCAA 2032

Qy 317 GGAACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGATTCGAGTATGG 376
Db 2033 GGAACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGATTCGAGTATGG 2092

Qy 377 CGGCGAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTCGCGCCCTGATGTC 436
Db 2093 CGGCGAGGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTCGCGCCCTGATGTC 2152

Qy 437 CACCGAGGCTCCGAGAACATCACTACCACTGCAAGAAACAGCGTGGCCCTACATGGACCA 496
Db 2153 CACCGAGGCTCCGAGAACATCACTACCACTGCAAGAAACAGCGTGGCCCTACATGGACCA 2212

Qy 497 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556
Db 2213 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 2272

Qy 557 CGCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGAGTCAAC 616
Db 2273 CGCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGAGTCAAC 2332

Qy 617 CGGAGCTGGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCCCTGCCAT 676
Db 2333 CGGAGCTGGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCCGCCCTGCCAT 2392

Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGGCCCGAGACAGGAATTCGGCTTCGACGTTGG 736
Db 2393 CATCGATGTGGCCCTTGGACGTTGGTGGCCCGAGACAGGAATTCGGCTTCGACGTTGG 2452

Qy 737 CCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 771
Db 2453 CCCTGTCTGCTTCTCTGTAAACTCCCTCCATCTAGA 2487
```

```

RESULT 9
US-10-177-293-66
; Sequence 66, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
```

APPLICANT: Glatt, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarpu, Manjula
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Bast Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Pusztai, Lajos
APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Mills, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
FILE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 66
LENGTH: 5921
TYPE: DNA
ORGANISM: Homo sapiens
US-10-177-293-66

Query Match 97.4%; Score 751; DB 6; Length 5921;
Best Local Similarity 100.0%; Pred. No. 9.9e-209;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGTTCTGTCGACCGTGCAGGTGGACACCACTTCAAGAGCCTGAG 76
Db 3776 TGATGCCAATGTGTTCTGTCGACCGTGCAGGTGGACACCACTTCAAGAGCCTGAG 3835
Qy 77 CCAGCAGATCGAAGCATCCGAGGCCAGAGGGAGCCGCAAGAACCCCGCCGACCTG 136
Db 3836 CCAGCAGATCGAAGCATCCGAGGCCAGAGGGAGCCGCAAGAACCCCGCCGACCTG 3895
Qy 137 CCGTGACCTCAAGATGTGCACCTCTGACTGGAAGAGTGGAGACTTCTGGATTGACCCCAA 196
Db 3896 CCGTGACCTCAAGATGTGCACCTCTGACTGGAAGAGTGGAGACTTCTGGATTGACCCCAA 3955
Qy 197 CCAAGGCTGCAACTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTGGAGCCTG 256
Db 3956 CCAAGGCTGCAACTGGATGCCATCAAAGTCTTCTGCAACATGGAGACTGGTGGAGCCTG 4015
Qy 257 CGTGATACCCCACTCAGCCAGTGTGGCCAGAGAGACTGGTATCATCAGCAAGAACCCCAA 316
Db 4016 CGTGATACCCCACTCAGCCAGTGTGGCCAGAGAGACTGGTATCATCAGCAAGAACCCCAA 4075
Qy 317 GGAACAAGAGGCATGTCTGGTTTGGCCGAGAGCATCACCGATGGATTCCAGTTCCAGTATGG 376
Db 4076 GGAACAAGAGGCATGTCTGGTTTGGCCGAGAGCATCACCGATGGATTCCAGTTCCAGTATGG 4135
Qy 377 CGGCGAGGGCTCGACCCCTGCCAGATGFGCCATCCAGCTGACCTTCTGCGCCTGATGTC 436
Db 4136 CGGCGAGGGCTCGACCCCTGCCAGATGFGCCATCCAGCTGACCTTCTGCGCCTGATGTC 4195

Qy 437 CACCGAGGCTTCCAGAACATCACCTACCACTGCAAGAAACAGCGTGGCTTACATGGAACA 496
Db 4196 CACCGAGGCTTCCAGAACATCACCTACCACTGCAAGAAACAGCGTGGCTTACATGGAACA 4255
Qy 497 GCAGACTGGAACTCAAGAAGCCCTGCTCTCAAGGGCTCAACGAGATCGAGATCCG 556
Db 4256 GCAGACTGGAACTCAAGAAGCCCTGCTCTCAAGGGCTCAACGAGATCGAGATCCG 4315
Qy 557 CGCCGAGGCAACAGCCGCTTCACTACAGCGTCACTGTCTGATGGTGCACGAGTCAAC 616
Db 4316 CGCCGAGGCAACAGCCGCTTCACTACAGCGTCACTGTCTGATGGTGCACGAGTCAAC 4375
Qy 617 CGGAGCCTGGGGCAACAGCAGTGAATTGAATACAAAACCAACAAGTCTCCGCGCTGCCCAT 676
Db 4376 CGGAGCCTGGGGCAACAGCAGTGAATTGAATACAAAACCAACAAGTCTCCGCGCTGCCCAT 4435
Qy 677 CATCGATGCGCCCTTGGAGCTTGGTGGCCCGAGACAGGAATTCGGCTTCGACGTTGG 736
Db 4436 CATCGATGCGCCCTTGGAGCTTGGTGGCCCGAGACAGGAATTCGGCTTCGACGTTGG 4495
Qy 737 CCTGTCTGCTTCTGTTAACTCCCTCCATC 767
Db 4496 CCTGTCTGCTTCTGTTAACTCCCTCCATC 4526

RESULT 10

US-10-301-822-29
Sequence 29, Application US/10301822
Publication No. US20030148410A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Berger, Allison
APPLICANT: Guillemette, Tracy L.
APPLICANT: Kamatkar, Shubhangi
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Thibodeau, Stephen N.
APPLICANT: Burgart, Lawrence J.
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF COLON CANCER
FILE REFERENCE: MPM01-029P2RNM
CURRENT APPLICATION NUMBER: US/10/301,822
CURRENT FILING DATE: 2002-11-21
PRIOR APPLICATION NUMBER: US 60/339,971
PRIOR FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: US 60/361,978
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/381,988
PRIOR FILING DATE: 2002-05-20
NUMBER OF SEQ ID NOS: 228
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 5921
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (120)...(4514)
US-10-301-822-29

Query Match 97.4%; Score 751; DB 6; Length 5921;
Best Local Similarity 100.0%; Pred. No. 9.9e-209;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGTTCTGTCGACCGTGCAGGTGGACCTTCAAGAGCCTGAG 76
Db 3776 TGATGCCAATGTGTTCTGTCGACCGTGCAGGTGGACCTTCAAGAGCCTGAG 3835
Qy 77 CCAGCAGATCGAAGCATCCGAGGCCAGAGGGAGCCGCAAGAACCCCGCCGACCTG 136
Db 3836 CCAGCAGATCGAAGCATCCGAGGCCAGAGGGAGCCGCAAGAACCCCGCCGACCTG 3895


```
QY 137 CCGTGACCTCAAGATGTCCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 196
Db |||||||
QY 3896 CCGTGACCTCAAGATGTCCTGACTGGAAGTGGAGAGTACTGGATTGACCCCAA 3955
Db |||||||
QY 197 CCAAGGCTGCAACCTGGATGTCATCAAAAGTCTTCTGCAACATGGAGAGTGGTGGAGACTG 256
Db |||||||
QY 3956 CCAAGGCTGCAACCTGGATGTCATCAAAAGTCTTCTGCAACATGGAGAGTGGTGGAGACTG 4015
QY 257 CGTGATACCCCACTCAGACCCAGTGTGGCCAGAGAACTGGTACATCAAGAGAAACCCCAA 316
Db |||||||
QY 4016 CGTGATACCCCACTCAGACCCAGTGTGGCCAGAGAACTGGTACATCAAGAGAAACCCCAA 4075
QY 317 GGACAGAGGAGTCTGTGTTCCGGGAGAGATGACCGATGATTCAGTTCCGAGTATGG 376
Db |||||||
QY 4076 GGACAGAGGAGTCTGTGTTCCGGGAGAGATGACCGATGATTCAGTTCCGAGTATGG 4135
QY 377 CGGCGAGGGCTCCAGACCTCGGATGTGGCCATCCAGCTGACCTTCTGCGCCCTGATGTC 436
Db |||||||
QY 4136 CGGCGAGGGCTCCAGACCTCGGATGTGGCCATCCAGCTGACCTTCTGCGCCCTGATGTC 4195
QY 437 CACGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCGTGGCTTACATGGACCA 496
Db |||||||
QY 4196 CACGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCGTGGCTTACATGGACCA 4255
QY 497 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGGCTCCAAAGAGATGAGATCCG 556
Db |||||||
QY 4256 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGGCTCCAAAGAGATGAGATCCG 4315
QY 557 CGGCGAGGGCAACAGCGCTTCACTACAGGTCACCTGTCGATGCTGCAAGTCAAC 616
Db |||||||
QY 4316 CGGCGAGGGCAACAGCGCTTCACTACAGGTCACCTGTCGATGCTGCAAGTCAAC 4375
QY 617 CGGAGCTGGGGCAAGACAGTGTGTAATACAAACCAACCAAGTCTCCCGCTGCCCAT 676
Db |||||||
QY 4376 CGGAGCTGGGGCAAGACAGTGTGTAATACAAACCAACCAAGTCTCCCGCTGCCCAT 4435
QY 677 CATCGATGTGGCCCTTTGAGAGTGTGGTGGCCCAAGACAGGAAATTCGGTTCGAGTGG 736
Db |||||||
QY 4436 CATCGATGTGGCCCTTTGAGAGTGTGGTGGCCCAAGACAGGAAATTCGGTTCGAGTGG 4495
QY 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
Db |||||||
QY 4496 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526
```

RESULT 11

```
US-10-764-425-18
; Sequence 18, Application US/10764425
; Publication No. US20040146921A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; APPLICANT: Taylor, Ian
; TITLE OF INVENTION: EXPRESSION PROFILES FOR COLON CANCER AND METHODS OF USE
; FILE REFERENCE: 5151
; CURRENT APPLICATION NUMBER: US/10764,425
; CURRENT FILING DATE: 2004-01-23
; PRIOR APPLICATION NUMBER: 60/442,582
; PRIOR FILING DATE: 2003-01-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-764-425-18
```

Query Match 97.4%; Score 751; DB 7; Length 5921;
Best Local Similarity 100.0%; Pred. No. 9.9e-209;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 17 TGATGCCAATGTGTTGTTGACCGTGACCTCGAGGTGGACACACCCCTCAAGAGCCTGAG 76
Db |||||||
QY 3776 TGATGCCAATGTGTTGTTGACCGTGACCTCGAGGTGGACACACCCCTCAAGAGCCTGAG 3835
Db |||||||
QY 77 CCAGCAGATCGAGAAATCCGGAGCCGAGAGGGAAGCCGCAAGAAACCCCGCCGACACTG 136
Db |||||||
QY 3836 CCAGCAGATCGAGAAATCCGGAGCCGAGAGGGAAGCCGCAAGAAACCCCGCCGACACTG 3895
QY 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db |||||||
QY 3896 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955
QY 197 CCAAGGCTGCAACCTGGATGTCATCAAAAGTCTTCTGCAACATGGAGAGTGGTGGAGACTG 256
Db |||||||
QY 3956 CCAAGGCTGCAACCTGGATGTCATCAAAAGTCTTCTGCAACATGGAGAGTGGTGGAGACTG 4015
QY 257 CGTGATACCCCACTCAGACCCAGTGTGGCCAGAGAACTGGTACATCAAGAGAAACCCCAA 316
Db |||||||
QY 4016 CGTGATACCCCACTCAGACCCAGTGTGGCCAGAGAACTGGTACATCAAGAGAAACCCCAA 4075
QY 317 GGACAGAGGAGTCTGTGTTCCGGGAGAGATGACCGATGATTCAGTTCCGAGTATGG 376
Db |||||||
QY 4076 GGACAGAGGAGTCTGTGTTCCGGGAGAGATGACCGATGATTCAGTTCCGAGTATGG 4135
QY 377 CGGCGAGGGCTCCAGACCTCGGATGTGGCCATCCAGCTGACCTTCTGCGCCCTGATGTC 436
Db |||||||
QY 4136 CGGCGAGGGCTCCAGACCTCGGATGTGGCCATCCAGCTGACCTTCTGCGCCCTGATGTC 4195
QY 437 CACGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCGTGGCTTACATGGACCA 496
Db |||||||
QY 4196 CACGAGGCTCCAGAACATCACTACCACTGCAAGAACAGCGTGGCTTACATGGACCA 4255
QY 497 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGGCTCCAAAGAGATGAGATCCG 556
Db |||||||
QY 4256 GCAGACTGGCAACTCAAGAGGCGCTGCTCTCAAGGGCTCCAAAGAGATGAGATCCG 4315
QY 557 CGGCGAGGGCAACAGCGCTTCACTACAGGTCACCTGTCGATGCTGCAAGTCAAC 616
Db |||||||
QY 4316 CGGCGAGGGCAACAGCGCTTCACTACAGGTCACCTGTCGATGCTGCAAGTCAAC 4375
QY 617 CGGAGCTGGGGCAAGACAGTGTGTAATACAAACCAACCAAGTCTCCCGCTGCCCAT 676
Db |||||||
QY 4376 CGGAGCTGGGGCAAGACAGTGTGTAATACAAACCAACCAAGTCTCCCGCTGCCCAT 4435
QY 677 CATCGATGTGGCCCTTTGAGAGTGTGGTGGCCCAAGACAGGAAATTCGGTTCGAGTGG 736
Db |||||||
QY 4436 CATCGATGTGGCCCTTTGAGAGTGTGGTGGCCCAAGACAGGAAATTCGGTTCGAGTGG 4495
QY 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
Db |||||||
QY 4496 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526
```

RESULT 12

```
US-10-734-564-12
; Sequence 12, Application US/10734564
; Publication No. US20040157278A1
; GENERAL INFORMATION:
; APPLICANT: Christopher C Burgess et al
; TITLE OF INVENTION: Detection Methods Using TIMP1
; FILE REFERENCE: 1657/2012
; CURRENT APPLICATION NUMBER: US/10734,564
; CURRENT FILING DATE: 2003-12-12
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-734-564-12
```

Query Match 97.4%; Score 751; DB 7; Length 5921;
Best Local Similarity 100.0%; Pred. No. 9.9e-209;


```
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 TGATGCCAATGTGTTGTCGACCGTGCACCTCGAGGTGGACACCACTCAAGAGCCTGAG 76
Db |||||||
QY 3776 TGATGCCAATGTGTTGTCGACCGTGCACCTCGAGGTGGACACCACTCAAGAGCCTGAG 3835
Db |||||||
QY 77 CCAGCAGATCGAGAACATCCGGAGCCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG 136
Db |||||||
QY 3836 CCAGCAGATCGAGAACATCCGGAGCCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG 3895
Db |||||||
QY 137 CCGTGACCTCAAGATGTGCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db |||||||
QY 3896 CCGTGACCTCAAGATGTGCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955
Db |||||||
QY 197 CCAAGGCTGCAACCTGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTG 4015
Db |||||||
QY 257 CGTGTACCCCACTCAGCCCACTGAGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db |||||||
QY 4016 CGTGTACCCCACTCAGCCCACTGAGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 4075
Db |||||||
QY 317 GGAACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
Db |||||||
QY 4076 GGAACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 4135
Db |||||||
QY 377 CGGCCAGGGCTCCGACCCCTGCGATGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 436
Db |||||||
QY 4136 CGGCCAGGGCTCCGACCCCTGCGATGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 4195
Db |||||||
QY 437 CACCGAGGCTCCAGAACATCACCTACCACTGCAAGAACAGGGTGGCCTACATGGACCA 496
Db |||||||
QY 4196 CACCGAGGCTCCAGAACATCACCTACCACTGCAAGAACAGGGTGGCCTACATGGACCA 4255
Db |||||||
QY 497 GCAGACTGGCAACTCAAGAGGCCCTGCTCTCAAGGGTCCAAAGAGATCGAGATCCG 556
Db |||||||
QY 4256 GCAGACTGGCAACTCAAGAGGCCCTGCTCTCAAGGGTCCAAAGAGATCGAGATCCG 4315
Db |||||||
QY 557 CGCCGAGGGCAAGCGCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGAGTCAAC 616
Db |||||||
QY 4316 CGCCGAGGGCAAGCGCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGAGTCAAC 4375
Db |||||||
QY 617 CGGAGCTGGGGCAAGACAGTGTGAATACAAACCCACCAAGTCTCTCCGCTGCCCAT 676
Db |||||||
QY 4376 CGGAGCTGGGGCAAGACAGTGTGAATACAAACCCACCAAGTCTCTCCGCTGCCCAT 4435
Db |||||||
QY 677 CATCGATGGCCCCCTTGAGCGTTGGTGGCCCGAGACCGAGAAATTCGGCTTCGACGTTGG 736
Db |||||||
QY 4436 CATCGATGGCCCCCTTGAGCGTTGGTGGCCCGAGACCGAGAAATTCGGCTTCGACGTTGG 4495
Db |||||||
QY 737 CCCTGTCTGCTCTCTGTAAACTCCCTCCATC 767
Db |||||||
QY 4496 CCCTGTCTGCTCTCTGTAAACTCCCTCCATC 4526
Db |||||||
```

```
RESULT 13
US-10-788-792-23
; Sequence 23, Application US/10788792
; Publication No. US20040191819A1
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Bigwood, Douglas
; APPLICANT: Eveleigh, Deepa
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT APPLICATION NUMBER: US/10/788,792
; PRIOR FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 5921
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-788-792-23
Query Match 97.4%; Score 751; DB 8; Length 5921;
Best Local Similarity 100.0%; Pred. No. 9.9e-209;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 TGATGCCAATGTGTTGTCGACCGTGCACCTCGAGGTGGACACCACTCAAGAGCCTGAG 76
Db |||||||
QY 3776 TGATGCCAATGTGTTGTCGACCGTGCACCTCGAGGTGGACACCACTCAAGAGCCTGAG 3835
Db |||||||
QY 77 CCAGCAGATCGAGAACATCCGGAGCCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG 136
Db |||||||
QY 3836 CCAGCAGATCGAGAACATCCGGAGCCCGAGGGAAGCCGCAAGAACCCCGCCGACCTG 3895
Db |||||||
QY 137 CCGTGACCTCAAGATGTGCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db |||||||
QY 3896 CCGTGACCTCAAGATGTGCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955
Db |||||||
QY 197 CCAAGGCTGCAACCTGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTG 256
Db |||||||
QY 3956 CCAAGGCTGCAACCTGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGGAGACTG 4015
Db |||||||
QY 257 CGTGTACCCCACTCAGCCCACTGAGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db |||||||
QY 4016 CGTGTACCCCACTCAGCCCACTGAGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 4075
Db |||||||
QY 317 GGAACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
Db |||||||
QY 4076 GGAACAAGAGGCATGTCTGGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 4135
Db |||||||
QY 377 CGGCCAGGGCTCCGACCCCTGCGATGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 436
Db |||||||
QY 4136 CGGCCAGGGCTCCGACCCCTGCGATGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 4195
Db |||||||
QY 437 CACCGAGGCTCCAGAACATCACCTACCACTGCAAGAACAGGGTGGCCTACATGGACCA 496
Db |||||||
QY 4196 CACCGAGGCTCCAGAACATCACCTACCACTGCAAGAACAGGGTGGCCTACATGGACCA 4255
Db |||||||
QY 497 GCAGACTGGCAACTCAAGAGGCCCTGCTCTCAAGGGTCCAAAGAGATCGAGATCCG 556
Db |||||||
QY 4256 GCAGACTGGCAACTCAAGAGGCCCTGCTCTCAAGGGTCCAAAGAGATCGAGATCCG 4315
Db |||||||
QY 557 CGCCGAGGGCAAGCGCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGAGTCAAC 616
Db |||||||
QY 4316 CGCCGAGGGCAAGCGCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGAGTCAAC 4375
Db |||||||
QY 617 CGGAGCTGGGGCAAGACAGTGTGAATACAAACCCACCAAGTCTCTCCGCTGCCCAT 676
Db |||||||
QY 4376 CGGAGCTGGGGCAAGACAGTGTGAATACAAACCCACCAAGTCTCTCCGCTGCCCAT 4435
Db |||||||
QY 677 CATCGATGGCCCCCTTGAGCGTTGGTGGCCCGAGACCGAGAAATTCGGCTTCGACGTTGG 736
Db |||||||
QY 4436 CATCGATGGCCCCCTTGAGCGTTGGTGGCCCGAGACCGAGAAATTCGGCTTCGACGTTGG 4495
Db |||||||
QY 737 CCCTGTCTGCTCTCTGTAAACTCCCTCCATC 767
Db |||||||
QY 4496 CCCTGTCTGCTCTCTGTAAACTCCCTCCATC 4526
Db |||||||
```

```
RESULT 14
US-10-723-860-2288
; Sequence 2288, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882.0193.NPUS01
; CURRENT APPLICATION NUMBER: US/10/723,860
```

; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429, 739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2288
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-723-860-2288

Query Match 97.4%; Score 751; DB 8; Length 5921;
Best Local Similarity 100.0%; Pred. No. 9.9e-209; Indels 0; Gaps 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCCCTCAAGAGCCTGAG 76
Db 3776 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCCCTCAAGAGCCTGAG 3835
Qy 77 CCAGCAGATCGAAGCATCCGAGCCCGAGGGAGCGCGAAGAACCCCGCCCGCAGCTG 136
Db 3836 CCAGCAGATCGAAGCATCCGAGCCCGAGGGAGCGCGAAGAACCCCGCCCGCAGCTG 3895
Qy 137 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGTACTGGATTGACCCCAA 196
Db 3896 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGTACTGGATTGACCCCAA 3955
Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
Db 3956 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 4015
Qy 257 CGTGTACCCCACTCAGGCCAGTGTGGCCAGAGAACTGTGTACATGACGAAGAACCCCAA 316
Db 4016 CGTGTACCCCACTCAGGCCAGTGTGGCCAGAGAACTGTGTACATGACGAAGAACCCCAA 4075
Qy 317 GGAACAAGGAGCATCTGTGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCCGAGTATGG 376
Db 4076 GGAACAAGGAGCATCTGTGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCCGAGTATGG 4135
Qy 377 CGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTCGCGCCTGATGTC 436
Db 4136 CGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTCGCGCCTGATGTC 4195
Qy 437 CACGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCGTGGCCTACATGGACCA 496
Db 4196 CACGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCGTGGCCTACATGGACCA 4255
Qy 497 GCAGACTGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCCGATCCG 556
Db 4256 GCAGACTGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCCGATCCG 4315
Qy 4316 CGCCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCCGATCCG 616
Db 4375 CGCCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCCGATCCG 4375
Qy 617 CGGAGCCTGGGGCAAGACAGTGTGTAATAAACAACCAAGTCTTCCCGCCTGCCCAT 676
Db 4376 CGGAGCCTGGGGCAAGACAGTGTGTAATAAACAACCAAGTCTTCCCGCCTGCCCAT 4435
Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGCCCGCAGACGAGGAATTCGGCTTCGACGTTGG 736
Db 4436 CATCGATGTGGCCCTTGGACGTTGGTGCCCGCAGACGAGGAATTCGGCTTCGACGTTGG 4495
Qy 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
Db 4496 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526

RESULT 15
US-10-956-157-340
; Sequence 340, Application US/10956157
; Publication No. US20050118625A1
; GENERAL INFORMATION:

; APPLICANT: Wyeth
; APPLICANT: Mounts, William
; TITLE OF INVENTION: NUCLEIC ACID ARRAYS FOR DETECTING GENE EXPRESSION ASSOCIATED WITH
; TITLE OF INVENTION: HUMAN OSTEOARTHRITIS AND HUMAN PROTEASES
; FILE REFERENCE: 031896-043000 (AM 101081)
; CURRENT APPLICATION NUMBER: US/10/956,157
; CURRENT FILING DATE: 2004-10-04
; NUMBER OF SEQ ID NOS: 319805
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 340
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-956-157-340

Query Match 97.4%; Score 751; DB 9; Length 5921;
Best Local Similarity 100.0%; Pred. No. 9.9e-209; Indels 0; Gaps 0;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCCCTCAAGAGCCTGAG 76
Db 3776 TGATGCCAATGTGGTTCGTGACCGTGACCTCGAGGTGGACACCCCTCAAGAGCCTGAG 3835
Qy 77 CCAGCAGATCGAAGCATCCGAGCCCGAGGGAGCGCGAAGAACCCCGCCCGCAGCTG 136
Db 3836 CCAGCAGATCGAAGCATCCGAGCCCGAGGGAGCGCGAAGAACCCCGCCCGCAGCTG 3895
Qy 137 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGTACTGGATTGACCCCAA 196
Db 3896 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGTGGAGTACTGGATTGACCCCAA 3955
Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
Db 3956 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 4015
Qy 257 CGTGTACCCCACTCAGGCCAGTGTGGCCAGAGAACTGTGTACATGACGAAGAACCCCAA 316
Db 4016 CGTGTACCCCACTCAGGCCAGTGTGGCCAGAGAACTGTGTACATGACGAAGAACCCCAA 4075
Qy 317 GGAACAAGGAGCATCTGTGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCCGAGTATGG 376
Db 4076 GGAACAAGGAGCATCTGTGTTCCGCGAGAGCATGACCGATGGATTCCAGTTCCGAGTATGG 4135
Qy 377 CGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTCGCGCCTGATGTC 436
Db 4136 CGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTCGCGCCTGATGTC 4195
Qy 437 CACGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCGTGGCCTACATGGACCA 496
Db 4196 CACGAGGCTCCGAGAACATCACTACCACTGCAAGAACAGCGTGGCCTACATGGACCA 4255
Qy 497 GCAGACTGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCCGATCCG 556
Db 4256 GCAGACTGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCCGATCCG 4315
Qy 4316 CGCCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCCGATCCG 616
Db 4375 CGCCGAGGGCAACAGCCGCTTCACTACAGCGTCACTGTGATGGCTGCAAGATCCGATCCG 4375
Qy 617 CGGAGCCTGGGGCAAGACAGTGTGTAATAAACAACCAAGTCTTCCCGCCTGCCCAT 676
Db 4376 CGGAGCCTGGGGCAAGACAGTGTGTAATAAACAACCAAGTCTTCCCGCCTGCCCAT 4435
Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGCCCGCAGACGAGGAATTCGGCTTCGACGTTGG 736
Db 4436 CATCGATGTGGCCCTTGGACGTTGGTGCCCGCAGACGAGGAATTCGGCTTCGACGTTGG 4495
Qy 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
Db 4496 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526

Search completed: April 27, 2006, 19:25:02

Job time : 699.19 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 18:58:56 ; Search time 530.452 Seconds
(without alignments)
5902.968 Million cell updates/sec

Title: US-10-677-877A-3

Perfect score: 771

Sequence: 1 agcttctacgtaagtctgat.....gtaaactccctccatctaga 771

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA New:

- 1: /SIDSS/ptodata/1/pubpna/US08 NEW PUB.seq.*
- 2: /SIDSS/ptodata/1/pubpna/US06 NEW PUB.seq.*
- 3: /SIDSS/ptodata/1/pubpna/US07 NEW PUB.seq.*
- 4: /SIDSS/ptodata/1/pubpna/US09 NEW PUB.seq.*
- 5: /SIDSS/ptodata/1/pubpna/US09 NEW PUB.seq.*
- 6: /SIDSS/ptodata/1/pubpna/US09 NEW PUB.seq.*
- 7: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 8: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 9: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 10: /SIDSS/ptodata/1/pubpna/US10 NEW PUB.seq.*
- 11: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*
- 12: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*
- 13: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*
- 14: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*
- 15: /SIDSS/ptodata/1/pubpna/US11 NEW PUB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	751	97.4	5921	14	US-11-186-284-29
2	751	97.4	6728	9	Sequence 29, Appl
3	751	97.4	6728	14	Sequence 138, Appl
4	751	97.4	6728	14	Sequence 27, Appl
5	746.2	96.8	4749	8	Sequence 1, Appli
6	746.2	96.8	4770	14	Sequence 244, Appl
7	746.2	96.8	4770	14	Sequence 7, Appli
8	746.2	96.8	4791	14	Sequence 479, Appl
9	745.8	96.7	6741	14	Sequence 105, Appl
10	745.8	96.7	6741	14	Sequence 106, Appl
11	743	96.4	4752	7	Sequence 583, Appl
12	743	96.4	4752	7	Sequence 1017, Ap
13	654.2	84.9	659	14	Sequence 1195, Ap
14	640.6	83.1	5721	7	Sequence 991, Appl
15	640.6	83.1	5721	7	Sequence 81, Appl
16	640.6	83.1	5721	7	Sequence 215, Appl
17	640.6	83.1	5721	7	Sequence 826, Appl
18	636.2	82.5	642	14	Sequence 969, Appl
					Sequence 966, Appl

19	634.2	82.3	639	14	US-11-108-172-1003	Sequence 1003, Ap
20	632.2	82.0	637	14	US-11-108-172-1018	Sequence 1018, Ap
21	631.6	81.9	650	14	US-11-108-172-967	Sequence 967, App
22	631.2	81.9	636	14	US-11-108-172-1005	Sequence 1005, Ap
23	631.2	81.9	636	14	US-11-108-172-1022	Sequence 1022, Ap
24	629.4	81.6	5868	14	US-11-136-527-3354	Sequence 3354, Ap
25	623.2	80.8	629	14	US-11-108-172-1006	Sequence 1006, Ap
26	618.2	80.2	623	14	US-11-108-172-1019	Sequence 1019, Ap
27	616.2	79.9	621	14	US-11-108-172-1014	Sequence 1014, Ap
28	615.2	79.8	622	14	US-11-108-172-994	Sequence 994, App
29	612.2	79.4	629	14	US-11-108-172-968	Sequence 968, App
30	608.2	78.9	613	14	US-11-108-172-1002	Sequence 1002, Ap
31	593.8	77.0	609	14	US-11-108-172-972	Sequence 972, App
32	569.8	73.9	575	14	US-11-108-172-1007	Sequence 1007, Ap
33	563.8	73.1	576	14	US-11-108-172-920	Sequence 920, App
c	551.2	71.5	556	14	US-11-108-172-985	Sequence 985, App
	426.4	55.3	5060	11	US-11-202-057-4	Sequence 2, Appli
	426.4	55.3	5060	11	US-11-202-057-6	Sequence 4, Appli
	416.4	50.4	4864	14	US-11-136-527-245	Sequence 245, App
	388.4	54.0	390	14	US-11-108-172-1001	Sequence 1001, App
c	332.2	43.1	5453	14	US-11-091-883-108	Sequence 108, App
	330.4	42.9	4474	11	US-11-036-196-1986	Sequence 1986, Ap
	330.4	42.9	5086	14	US-11-186-284-30	Sequence 30, Appl
44	330.4	42.9	5086	14	US-11-091-883-109	Sequence 109, App
45	324	42.0	4101	8	US-10-821-234-579	Sequence 579, App

ALIGNMENTS

RESULT 1

US-11-186-284-29
; Sequence 29, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: BURGART, LAWRENCE J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: NPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 5921
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)...(4514)
US-11-186-284-29

Query Match 97.4%; Score 751; DB 14; Length 5921;
Best Local Similarity 100.0%; Pred. No. 5.5e-191;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

17 TGATGCCAATGTGTTGTCACCGTGTGACCTCGAGGTGGACACACACCTCTCAAGAGCCTGAG 76
Db
3776 TGATGCCAATGTGTTGTCACCGTGTGACCTCGAGGTGGACACACACCTCTCAAGAGCCTGAG 3835
Qy 77 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG 136
Db 3836 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG 3895
Qy 137 CCGTGACCTCAAGATGTCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3896 CCGTGACCTCAAGATGTCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955
Qy 197 CCAAGGCTGCAACTGGATGSCCATCAAGTCTTCTGCAACATGGAGACTGCTGAGACCTG 256
Db 3956 CCAAGGCTGCAACTGGATGSCCATCAAGTCTTCTGCAACATGGAGACTGCTGAGACCTG 4015
Qy 257 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAAGCTGGTACATCAGCAAGAACCCCAA 316
Db 4016 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAAGCTGGTACATCAGCAAGAACCCCAA 4075
Qy 317 GGAACAGAGCATGTCTGGTTGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
Db 4076 GGAACAGAGCATGTCTGGTTGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 4135
Qy 377 CGGCGAGGGCTCCGACCCCTCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 436
Db 4136 CGGCGAGGGCTCCGACCCCTCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 4195
Qy 437 CACCGAGGGCTCCGAGAACATCACTACCTACCTGCAAGAACAGCGTGGCTTACATGGAACCA 496
Db 4196 CACCGAGGGCTCCGAGAACATCACTACCTACCTGCAAGAACAGCGTGGCTTACATGGAACCA 4255
Qy 497 GCGAGCTGGCAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGATCGAGTCCG 556
Db 4256 GCGAGCTGGCAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGATCGAGTCCG 4315
Qy 557 CGCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTCGATGGCTGCAAGAGTCCAGTCCG 616
Db 4316 CGCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTCGATGGCTGCAAGAGTCCAGTCCG 4375
Qy 617 CGGAGCTGGGGCAAGACAGTGAATTGAATACAAACCAACCAAGTCTTCCGCGCTGCCCAT 676
Db 4376 CGGAGCTGGGGCAAGACAGTGAATTGAATACAAACCAACCAAGTCTTCCGCGCTGCCCAT 4435
Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGGCCCAAGACAGGAATTCGGCTTCGACGTTGG 736
Db 4436 CATCGATGTGGCCCTTGGACGTTGGTGGCCCAAGACAGGAATTCGGCTTCGACGTTGG 4495
Qy 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
Db 4496 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526

RESULT 2
US-10-501-035-138
; Sequence 138, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; TITLE OF INVENTION: AND/OR PROTEIN TYROSINE KINASE PATHWAYS
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501.035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 138
; LENGTH: 6728
; TYPE: DNA

; ORGANISM: Homo sapiens
US-10-501-035-138
Query Match 97.4%; Score 751; DB 9; Length 6728;
Best Local Similarity 100.0%; Pred. No. 5.6e-191; Indels 0; Gaps 0;
Matches 751; Conservative 0; Mismatches 0;
Qy 17 TGATGCCAATGTGTTGTCACCGTGTGACCTCGAGGTGGACACACCTCTCAAGAGCCTGAG 76
Db 3776 TGATGCCAATGTGTTGTCACCGTGTGACCTCGAGGTGGACACACCTCTCAAGAGCCTGAG 3835
Qy 77 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG 136
Db 3836 CCAGCAGATCGAGAACATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGACCTG 3895
Qy 137 CCGTGACCTCAAGATGTCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3896 CCGTGACCTCAAGATGTCACCTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955
Qy 197 CCAAGGCTGCAACTGGATGSCCATCAAGTCTTCTGCAACATGGAGACTGCTGAGACCTG 256
Db 3956 CCAAGGCTGCAACTGGATGSCCATCAAGTCTTCTGCAACATGGAGACTGCTGAGACCTG 4015
Qy 257 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAAGCTGGTACATCAGCAAGAACCCCAA 316
Db 4016 CGTGTACCCCACTCAGCCCAAGTGTGGCCCAAGAAAGCTGGTACATCAGCAAGAACCCCAA 4075
Qy 317 GGAACAGAGCATGTCTGGTTGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 376
Db 4076 GGAACAGAGCATGTCTGGTTGGCGAGAGCATGACCGATGGATTCCAGTTCGAGTATGG 4135
Qy 377 CGGCGAGGGCTCCGACCCCTCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 436
Db 4136 CGGCGAGGGCTCCGACCCCTCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 4195
Qy 437 CACCGAGGGCTCCGAGAACATCACTACCTACCTGCAAGAACAGCGTGGCTTACATGGAACCA 496
Db 4196 CACCGAGGGCTCCGAGAACATCACTACCTACCTGCAAGAACAGCGTGGCTTACATGGAACCA 4255
Qy 497 GCGAGCTGGCAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGATCGAGTCCG 556
Db 4256 GCGAGCTGGCAACCTCAAGAGGCCCTGCTCTCAAGGGCTCCAAAGAGATCGAGTCCG 4315
Qy 557 CGCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTCGATGGCTGCAAGAGTCCAGTCCG 616
Db 4316 CGCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTCGATGGCTGCAAGAGTCCAGTCCG 4375
Qy 617 CGGAGCTGGGGCAAGACAGTGAATTGAATACAAACCAACCAAGTCTTCCGCGCTGCCCAT 676
Db 4376 CGGAGCTGGGGCAAGACAGTGAATTGAATACAAACCAACCAAGTCTTCCGCGCTGCCCAT 4435
Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGGCCCAAGACAGGAATTCGGCTTCGACGTTGG 736
Db 4436 CATCGATGTGGCCCTTGGACGTTGGTGGCCCAAGACAGGAATTCGGCTTCGACGTTGG 4495
Qy 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
Db 4496 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526

RESULT 3
US-11-186-284-27
; Sequence 27, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.

```
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)...(4514)
US-11-186-284-27

Query Match          97.4%; Score 751; DB 14; Length 6728;
Best Local Similarity 100.0%; Pred. No. 5.6e-191;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 TGATGCCAATGTGGTTCGTGACCGTGTGACCGTGTGAGTGGACACACCCCTCAAGAGCCCTGAG 76
Db 3776 TGATGCCAATGTGGTTCGTGACCGTGTGACCGTGTGAGTGGACACACCCCTCAAGAGCCCTGAG 3835

Qy 77 CCAGCAGATCGAGAACATCCGAGCCCGAGAGCGGCAAGAACCCCGCCGCGCACCTG 136
Db 3836 CCAGCAGATCGAGAACATCCGAGCCCGAGAGCGGCAAGAACCCCGCCGCGCACCTG 3895

Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3896 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955

Qy 197 CCAAGGCTGCAACTCGATGCTCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
Db 3956 CCAAGGCTGCAACTCGATGCTCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 4015

Qy 257 CGTGATACCCCACTCAGCCCACTGAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 4016 CGTGATACCCCACTCAGCCCACTGAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 4075

Qy 317 GGAACAAGAGGATGTCTGGTTCCGCGAGAGCATGACCCGATGGATTCCAGTTCGAGTATGG 376
Db 4076 GGAACAAGAGGATGTCTGGTTCCGCGAGAGCATGACCCGATGGATTCCAGTTCGAGTATGG 4135

Qy 377 CGGCCAGGGCTCCGACCCCTCGGATGTGGCCATCCAGTCACTTCTCGCTGCGCTGATGTC 436
Db 4136 CGGCCAGGGCTCCGACCCCTCGGATGTGGCCATCCAGTCACTTCTCGCTGCGCTGATGTC 4195

Qy 437 CACGAGGCTCCAGAGAGGCCCTCTGCTCTCAAGAGGCTCCAAAGGCTTACATGGACCA 496
Db 4196 CACGAGGCTCCAGAGAGGCCCTCTGCTCTCAAGAGGCTCCAAAGGCTTACATGGACCA 4255

Qy 497 GCAGACTGGCAACTCAAGAGGCCCTCTGCTCTCAAGAGGCTCCAAAGGATCGAGATCCG 556
Db 4256 GCAGACTGGCAACTCAAGAGGCCCTCTGCTCTCAAGAGGCTCCAAAGGATCGAGATCCG 4315

Qy 557 GCGCAGGGCAACAGCCGCTTACCTACAGGTCACCTGTGATGGCTGCAAGATCAAC 616
Db 4316 GCGCAGGGCAACAGCCGCTTACCTACAGGTCACCTGTGATGGCTGCAAGATCAAC 4375

Qy 617 CGGAGCTCGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCCGCTGCCCAT 676
Db 4376 CGGAGCTCGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCCGCTGCCCAT 4435
```

```
Qy 677 CATCGATGTGCCCCCTTGGACGTTGGTGGCCCCCAGACCAAGAAATTCGGCTTCGACGTTGG 736
Db 4436 CATCGATGTGCCCCCTTGGACGTTGGTGGCCCCCAGACCAAGAAATTCGGCTTCGACGTTGG 4495

Qy 737 CCCTGTCTGCTTCCCTGTAAACTCCCTCCATC 767
Db 4496 CCCTGTCTGCTTCCCTGTAAACTCCCTCCATC 4526

RESULT 4
US-11-021-603-1
; Sequence 1, Application US/11021603
; Publication No. US20060003954A1
; GENERAL INFORMATION:
; APPLICANT: Berl, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/11/021,603
; CURRENT FILING DATE: 2004-12-21
; PRIOR APPLICATION NUMBER: US/10/149,352
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 1
; LENGTH: 6728
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (120)..(4511)
; NAME/KEY: sig peptide
; LOCATION: (120)..(185)
US-11-021-603-1

Query Match          97.4%; Score 751; DB 14; Length 6728;
Best Local Similarity 100.0%; Pred. No. 5.6e-191;
Matches 751; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 17 TGATGCCAATGTGGTTCGTGACCGTGTGACCGTGTGAGTGGACACACCCCTCAAGAGCCCTGAG 76
Db 3776 TGATGCCAATGTGGTTCGTGACCGTGTGACCGTGTGAGTGGACACACCCCTCAAGAGCCCTGAG 3835

Qy 77 CCAGCAGATCGAGAACATCCGAGCCCGAGAGCGGCAAGAACCCCGCCGCGCACCTG 136
Db 3836 CCAGCAGATCGAGAACATCCGAGCCCGAGAGCGGCAAGAACCCCGCCGCGCACCTG 3895

Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3896 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3955

Qy 197 CCAAGGCTGCAACTCGATGCTCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
Db 3956 CCAAGGCTGCAACTCGATGCTCAAAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 4015

Qy 257 CGTGATACCCCACTCAGCCCACTGAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 4016 CGTGATACCCCACTCAGCCCACTGAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 4075

Qy 317 GGAACAAGAGGATGTCTGGTTCCGCGAGAGCATGACCCGATGGATTCCAGTTCGAGTATGG 376
Db 4076 GGAACAAGAGGATGTCTGGTTCCGCGAGAGCATGACCCGATGGATTCCAGTTCGAGTATGG 4135

Qy 377 CGGCCAGGGCTCCGACCCCTCGGATGTGGCCATCCAGTCACTTCTCGCTGCGCTGATGTC 436
Db 4136 CGGCCAGGGCTCCGACCCCTCGGATGTGGCCATCCAGTCACTTCTCGCTGCGCTGATGTC 4195

Qy 437 CACGAGGCTCCAGAGAGGCCCTCTGCTCTCAAGAGGCTCCAAAGGCTTACATGGACCA 496
Db 4196 CACGAGGCTCCAGAGAGGCCCTCTGCTCTCAAGAGGCTCCAAAGGCTTACATGGACCA 4255

Qy 497 GCAGACTGGCAACTCAAGAGGCCCTCTGCTCTCAAGAGGCTCCAAAGGATCGAGATCCG 556
Db 4256 GCAGACTGGCAACTCAAGAGGCCCTCTGCTCTCAAGAGGCTCCAAAGGATCGAGATCCG 4315

Qy 557 GCGCAGGGCAACAGCCGCTTACCTACAGGTCACCTGTGATGGCTGCAAGATCAAC 616
Db 4316 GCGCAGGGCAACAGCCGCTTACCTACAGGTCACCTGTGATGGCTGCAAGATCAAC 4375

Qy 617 CGGAGCTCGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCCGCTGCCCAT 676
Db 4376 CGGAGCTCGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCCGCTGCCCAT 4435
```

Db 4196 CACCGAGGCTCCAGAACATACCTACCACTGCAAGAAACAGCGTGGCCCTACATGGACCA 4255
Qy 497 GCAGACTGGCAACTCAAGAGGCTCCCTCAAGGGCTCCCAAGAGATCGAGATCCG 556
Db 4256 GCAGACTGGCAACTCAAGAGGCTCCCTCAAGGGCTCCCAAGAGATCGAGATCCG 4315
Qy 557 CGCGAGGGCAACAGCGCTTACCTACACGCTCACTGTGCGTGGCTGCAAGAGTCAAC 616
Db 4316 CGCGAGGGCAACAGCGCTTACCTACACGCTCACTGTGCGTGGCTGCAAGAGTCAAC 4375
Qy 617 CGAGGCTGGGGCAAGACAGTGAATGAATACCAACCAAGTCTCCCGCTGGCCAT 676
Db 4376 CGAGGCTGGGGCAAGACAGTGAATGAATACCAACCAAGTCTCCCGCTGGCCAT 4435
Qy 677 CATCGATGGGCCCCCTTGGACGTTGGTCCCGACAGACCAAGAAATCGGCTTCGACGTTGG 736
Db 4436 CATCGATGGGCCCCCTTGGACGTTGGTCCCGACAGACCAAGAAATCGGCTTCGACGTTGG 4495
Qy 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
Db 4496 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526
RESULT 5
US-10-821-234-244
; Sequence 244, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: PCT_SEQ_genes Version 1.0
; SEQ ID NO 244
; LENGTH: 4749
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-821-234-244
Query Match 96.8%; Score 746.2; DB 8; Length 4749;
Best Local Similarity 99.6%; Pred. No. 1e-189;
Matches 748; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 17 TGAATGCAATGTGGTTGCTGACCGTGACCTCGAGGTGGACACCACTCAAGAGCCTGAG 76
Db 3776 TGAATGCAATGTGGTTGCTGACCGTGACCTCGAGGTGGACACCACTCAAGAGCCTGAG 3835
Qy 77 CCAGCAGATCGAAGCAATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTG 136
Db 3836 CCAGCAGATCGAAGCAATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTG 3895
Qy 137 CCAGCAGATCGAAGCAATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTG 196
Db 3896 CCAGCAGATCGAAGCAATCCGAGGCCAGAGGGAAGCCGCAAGAACCCCGCCGCACTG 3955
Qy 197 CCAGGCTGCAACTGGATGCCATCAAGCTTCTGCAACATGGAGACTGGTGGACCTG 256
Db 3956 CCAGGCTGCAACTGGATGCCATCAAGCTTCTGCAACATGGAGACTGGTGGACCTG 4015
Qy 257 CGTGTACCCCACTCAGCGCCAGTGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA 316
Db 4016 CGTGTACCCCACTCAGCGCCAGTGTGGCCAGAGAACTGGTATCATCAGCAAGAACCCCAA 4075
Qy 317 GGAACAAGAGCAATGTCTGGTTGGCGAGAGCATGACCGATGGATTCAGTTTCGAGTATGG 376
Db 4076 GGAACAAGAGCAATGTCTGGTTGGCGAGAGCATGACCGATGGATTCAGTTTCGAGTATGG 4135

Qy 377 CGGCCAGGGCTCCGAGCCCTGCGATGTGGCCATTCAGCTGACCTTCTCGCCTGATGTC 436
Db 4136 CGGCCAGGGCTCCGAGCCCTGCGATGTGGCCATTCAGCTGACCTTCTCGCCTGATGTC 4195
Qy 437 CACCGAGGCTCCAGAAATACCTACCATGCAAGAAACAGCGTGGCCCTACATGGACCA 496
Db 4196 CACCGAGGCTCCAGAAATACCTACCATGCAAGAAACAGCGTGGCCCTACATGGACCA 4255
Qy 497 GCAGACTGGCAACTCAAGAGGCTCCCTCAAGGGCTCCCAAGAGATCGAGATCCG 556
Db 4256 GCAGACTGGCAACTCAAGAGGCTCCCTCAAGGGCTCCCAAGAGATCGAGATCCG 4315
Qy 557 CGCGAGGGCAACAGCGCTTACCTACACGCTCACTGTGCGTGGCTGCAAGAGTCAAC 616
Db 4316 CGCGAGGGCAACAGCGCTTACCTACACGCTCACTGTGCGTGGCTGCAAGAGTCAAC 4375
Qy 617 CGAGGCTGGGGCAAGACAGTGAATGAATACCAACCAAGTCTCCCGCTGGCCAT 676
Db 4376 CGAGGCTGGGGCAAGACAGTGAATGAATACCAACCAAGTCTCCCGCTGGCCAT 4435
Qy 677 CATCGATGGGCCCCCTTGGACGTTGGTCCCGACAGACCAAGAAATCGGCTTCGACGTTGG 736
Db 4436 CATCGATGGGCCCCCTTGGACGTTGGTCCCGACAGACCAAGAAATCGGCTTCGACGTTGG 4495
Qy 737 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 767
Db 4496 CCCTGTCTGCTTCTGTAAACTCCCTCCATC 4526
RESULT 6
US-11-000-463-7
; Sequence 7, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785C1P4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR FILING DATE: 2004-11-29
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 4770
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140)..(4531)

US-11-000-463-7

Query Match 96.8%; Score 746.2; DB 14; Length 4770;
Best Local Similarity 99.6%; Pred. No. 1e-189;
Matches 748; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy 17 TGATGCCAATGTGTTCTGTCGACCGTGCACCTCGAGTGGACACACCTCAAGACCTGAG 76
Db 3796 TGATGCCAATGTGTTCTGTCGACCGTGCACCTCGAGTGGACACACCTCAAGACCTGAG 3855
Qy 77 CCACGAGATCCAGAACATCCGAGACCCAGAGGGAGCCGCAAGAACCCCGCCGACCTG 136
Db 3856 CCACGAGATCCAGAACATCCGAGACCCAGAGGGAGCCGCAAGAACCCCGCCGACCTG 3915
Qy 137 CCGTGAACCTCAAGATGTGCCACTCTGTGCTGAAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3916 CCGTGAACCTCAAGATGTGCCACTCTGTGCTGAAAGAGTGGAGAGTACTGGATTGACCCCAA 3975
Qy 197 CCAGGCTGCAACCTGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
Db 3976 CCAGGCTGCAACCTGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 4035
Qy 257 CGTGTACCCCACTCAGCCCACTGTGGCCGAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 4036 CGTGTACCCCACTCAGCCCACTGTGGCCGAGAGAACTGGTACATCAGCAAGAACCCCAA 4095
Qy 317 GGACAAGAGGATGTCTGGTTGCGGAGAGATGACCGATGGATTCCAGTTGAGTATGG 376
Db 4096 GGACAAGAGGATGTCTGGTTGCGGAGAGATGACCGATGGATTCCAGTTGAGTATGG 4155
Qy 377 CGGCGAGGGCTCCAGACCTGCGATGTGGCCATCCAGTGCATCTCTGCGCCTGATGTC 436
Db 4156 CGGCGAGGGCTCCAGACCTGCGATGTGGCCATCCAGTGCATCTCTGCGCCTGATGTC 4215
Qy 437 CACGAGCGCTCCAGAACATCACTTACCCTGCAAGAACAGCGTGGCTTACATGGACCA 496
Db 4216 CACGAGCGCTCCAGAACATCACTTACCCTGCAAGAACAGCGTGGCTTACATGGACCA 4275
Qy 497 GCAGACTGGCAACTCTCAAGAGGCGCTCTCTCTCAAGGGCTTCCAAAGAGATCGAGATCCG 556
Db 4276 GCAGACTGGCAACTCTCAAGAGGCGCTCTCTCTCAAGGGCTTCCAAAGAGATCGAGATCCG 4335
Qy 557 CGCGAGGGCAACAGCCGCTTCACTACAGGTCATCTGTGATGGCTGCAAGATCCAC 616
Db 4336 CGCGAGGGCAACAGCCGCTTCACTACAGGTCATCTGTGATGGCTGCAAGATCCAC 4395
Qy 617 CGGAGCTTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCCCGCCTGCCAT 676
Db 4396 CGGAGCTTGGGGCAAGACAGTGTGAATCAAAACCAACCAAGTCTCCCGCCTGCCAT 4455
Qy 677 CATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCAGACCAAGAAATTCGACGTTGG 736
Db 4456 CATCGATGTGGCCCCCTTGGACGTTGGTGGCCCCAGACCAAGAAATTCGACGTTGG 4515
Qy 737 CCTGTCTGCTCTCTGTAACCTCCTCATC 767
Db 4516 CCTGTCTGCTCTCTGTAACCTCCTCATC 4546
```

RESULT 7

US-11-000-463-479
; Sequence 479, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chonghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping

; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785C1P4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 479
; LENGTH: 4791
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(4791)
; OTHER INFORMATION: n = a,t,c or g
US-11-000-463-479

Query Match 96.8%; Score 746.2; DB 14; Length 4791;
Best Local Similarity 99.6%; Pred. No. 1e-189;
Matches 748; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy 17 TGATGCCAATGTGTTCTGTCGACCGTGCACCTCGAGTGGACACACCTCAAGACCTGAG 76
Db 3816 TGATGCCAATGTGTTCTGTCGACCGTGCACCTCGAGTGGACACACCTCAAGACCTGAG 3875
Qy 77 CCACGAGATCCAGAACATCCGAGACCCAGAGGGAGCCGCAAGAACCCCGCCGACCTG 136
Db 3876 CCACGAGATCCAGAACATCCGAGACCCAGAGGGAGCCGCAAGAACCCCGCCGACCTG 3935
Qy 137 CCGTGAACCTCAAGATGTGCCACTCTGTGCTGAAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3936 CCGTGAACCTCAAGATGTGCCACTCTGTGCTGAAAGAGTGGAGAGTACTGGATTGACCCCAA 3995
Qy 197 CCAGGCTGCAACCTGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 256
Db 3996 CCAGGCTGCAACCTGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGAGACTG 4055
Qy 257 CGTGTACCCCACTCAGCCCACTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 4056 CGTGTACCCCACTCAGCCCACTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 4115
Qy 317 GGACAAGAGGATGTCTGGTTGCGGAGAGCATCAACCGATGGATTCCAGTTGAGTATGG 376
Db 4116 GGACAAGAGGATGTCTGGTTGCGGAGAGCATCAACCGATGGATTCCAGTTGAGTATGG 4175
Qy 377 CGGCGAGGGCTCCAGACCTGCGATGGCCATCCAGTGCATCTCTGCGCCTGATGTC 436
Db 4176 CGGCGAGGGCTCCAGACCTGCGATGGCCATCCAGTGCATCTCTGCGCCTGATGTC 4235
Qy 437 CACGAGGGCTCCAGAACATCACTTACCCTGCAAGAACAGCGTGGCTTACATGGACCA 496
Db 4236 CACGAGGGCTCCAGAACATCACTTACCCTGCAAGAACAGCGTGGCTTACATGGACCA 4295
Qy 497 GCAGACTGGCAACTCTCAAGAGGCGCTCTCTCAAGGGCTTCCAAAGAGATCGAGATCCG 556
Db 4296 GCAGACTGGCAACTCTCAAGAGGCGCTCTCTCTCAAGGGCTTCCAAAGAGATCGAGATCCG 4355
Qy 557 CGCGAGGGCAACAGCCGCTTCACTTACAGCGTCACTGTGATGGCTGCAAGATCCAC 616
```

Db 4356 CGCGAGGGCAACAGCGCTTACCTACAGCGTCACTGTGATGGCTGCACGAGTCACAC 4415
Qy 617 CGAGAGCTGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCGGCTGCCCAT 676
Db 4416 CGGAGCTGGGCAAGACAGTGAATGAATACAAACCAACCAAGTCTCCGGCTGCCCAT 4475
Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGCCCGACAGACCAAGAAATCGGCTTCGACGTTGG 736
Db 4476 CATCGATGTGGCCCTTGGACGTTGGTGCCCGACAGACCAAGAAATCGGCTTCGACGTTGG 4535
Qy 737 CCTGTGTCTTCTGTAAACTCCCTCCATC 767
Db 4536 CCTGTGTCTTCTGTAAACTCCCTCCATC 4566

RESULT 8
US-11-091-883-105/c
; Sequence 105, Application US/11091883
; Publication No. US20060024693A1
; GENERAL INFORMATION:
; APPLICANT: CIBELLI, JOSE
; APPLICANT: FERNANDEZ, EMILIO O.
; APPLICANT: JORDAO DE MEGALHAES, GUILHERME
; APPLICANT: KOCABAS, ARTIF
; APPLICANT: CROSBY, JAVIER A.
; TITLE OF INVENTION: IDENTIFICATION OF GENES OR POLYPEPTIDES THE EXPRESSION OF WHICH
; TITLE OF INVENTION: CORRELATES TO FERTILITY, OVARIAN FUNCTION AND/OR FETAL/NEBORN
; TITLE OF INVENTION: VIABILITY
; FILE REFERENCE: 53942US
; CURRENT APPLICATION NUMBER: US/11/091,883
; CURRENT FILING DATE: 2005-03-29
; PRIOR APPLICATION NUMBER: 60/556,875
; PRIOR FILING DATE: 2004-03-29
; NUMBER OF SEQ ID NOS: 513
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 105
; LENGTH: 6741
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (758)..(758)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (760)..(760)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (762)..(766)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (768)..(775)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (777)..(782)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (810)..(810)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (812)..(812)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (928)..(928)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:

; NAME/KEY: modified_base
; LOCATION: (1203)..(1203)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1207)..(1207)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1420)..(1424)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1485)..(1485)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1501)..(1501)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1542)..(1545)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1692)..(1692)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1841)..(1841)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1892)..(1892)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1931)..(1938)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1940)..(1948)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2009)..(2009)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2871)..(2871)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2873)..(2873)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (3020)..(3020)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (3380)..(3380)
; OTHER INFORMATION: a, c, g, or t
US-11-091-883-105

Query Match 96.7%; Score 745.8; DB 14; Length 6741;
Best Local Similarity 99.5%; Pred. No. 1.4e-189;
Matches 747; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 17 TGATGCAATGTGTTCTGTGACCGTGCACCTCGAGGTGGACCAACCCCTCAAGAGCCTGAG 76
Db 2966 TGAIGCAATGTGTTCTGTGACCGTGCACCTCGAGGTGGACCAACCCCTCAAGAGCCTGAG 2907

```
Qy 77 CCAGCAGATCAGAAACATCCGAGCCAGAGGGAGCGCAAGAACCCCGCCCGCACCTG 136
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2906 CCAGCAGATCAGAAACATCCGAGCCAGAGGGAGCGCAAGAACCCCGCCCGCACCTG 2847
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 137 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2846 CCGTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 2787
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 197 CCAAGGCTGCAACCTGATGTCATCAAGTCTTCTGCAACATGAGAGTGGTGAAGCTG 256
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2786 CCAAGGCTGCAACCTGATGTCATCAAGTCTTCTGCAACATGAGAGTGGTGAAGCTG 2727
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 257 CGTGATCCCACTCAGCCCACTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2726 CGTGATCCCACTCAGCCCACTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 2667
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 317 GGACAAGAGGCAATGTCTGTTTCGGCGAGAGCATGACCGATGGATTCCAGTTTCGAGTATGG 376
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2666 GGACAAGAGGCAATGTCTGTTTCGGCGAGAGCATGACCGATGGATTCCAGTTTCGAGTATGG 2607
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 377 CGGCGAGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTTGGCCCTGATGTC 436
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2606 CGGCGAGGCTCCGACCCCTGCGATGTGGCCATCCAGCTGACCTTCTTGGCCCTGATGTC 2547
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 437 CACGAGGCTCCGACCAACATCCTACCTACCACTGCAAGACAGCGTGGCTTACATGGACCA 496
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2546 CACGAGGCTCCGACCAACATCCTACCTACCACTGCAAGACAGCGTGGCTTACATGGACCA 2487
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 497 GCAGACTGGCAACTCTCAAGAGGCGCTGCTCTCAAGGGCTTCCAAACGAGATCGAGATCCG 556
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2486 GCAGACTGGCAACTCTCAAGAGGCGCTGCTCTCAAGGGCTTCCAAACGAGATCGAGATCCG 2427
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 557 CGCGAGGCGCAACGCGCTTCACTACAGCGTCACCTGTCGATGCTGCGATGCTGCGATGCTCAC 616
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2426 CGCGAGGCGCAACGCGCTTCACTACAGCGTCACCTGTCGATGCTGCGATGCTGCGATGCTCAC 2367
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 617 CGGAGCTTGGGGCAAGACAGTGTGATTAACAAACACCAAGTCTCCCGCTTGGCCCAT 676
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2366 CGGAGCTTGGGGCAAGACAGTGTGATTAACAAACACCAAGTCTCCCGCTTGGCCCAT 2307
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 677 CATCGATGTGCCCCCTTGGAGCTTGGTGGCCCGACAGACAGGAATTCGGCTTCCAGCTGG 736
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2306 CATCGATGTGCCCCCTTGGAGCTTGGTGGCCCGACAGACAGGAATTCGGCTTCCAGCTGG 2247
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 737 CCTGTCTGCTTCTGTAACCTCCCTCCATC 767
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2246 CCTGTCTGCTTCTGTAACCTCCCTCCATC 2216
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 9
US-11-091-883-106/c
; Sequence 106, Application US/11091883
; Publication No. US20060024693A1
; GENERAL INFORMATION:
; APPLICANT: CIBELLI, JOSE
; APPLICANT: FERNANDEZ, EMILIO O.
; APPLICANT: JORDAO DE MEGALHAES, GUILHERME
; APPLICANT: KOCABAS, ARIF
; APPLICANT: CROSBY, JAVIER A.
; TITLE OF INVENTION: IDENTIFICATION OF GENES OR POLYPEPTIDES THE EXPRESSION OF WHICH
; TITLE OF INVENTION: CORRELATES TO FERTILITY, OVARIAN FUNCTION AND/OR FETAL/NEWBORN
; TITLE OF INVENTION: VIABILITY
; FILE REFERENCE: 53942US
; CURRENT APPLICATION NUMBER: US/11/091.883
; CURRENT FILING DATE: 2005-03-29
; PRIOR APPLICATION NUMBER: 60/556,875
; PRIOR FILING DATE: 2004-03-29
; NUMBER OF SEQ ID NOS: 513
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 106
; LENGTH: 6741
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (758)..(758)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (760)..(760)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (762)..(766)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (768)..(775)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (777)..(782)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (810)..(810)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (812)..(812)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (928)..(928)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1203)..(1203)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1207)..(1207)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1420)..(1424)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1485)..(1485)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1501)..(1501)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1542)..(1545)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1692)..(1692)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1841)..(1841)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1892)..(1892)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (1911)..(1938)
; OTHER INFORMATION: a, c, g, or t
; FEATURE:
```


Db 4373 CGGAGCCTGGGGCAAGACAGTGAATTAACAACCAACCAAGACCTCCCGCCTGGCGAT 4432
Qy 677 CATGATGTGGCCCTTGGAGCTGTGGTGGCCCAAGACCAAGAAATTCGGCTTCGACGTTGG 736
Db 4433 CATGATGTGGCCCTTGGAGCTGTGGTGGCCCAAGACCAAGAAATTCGGCTTCGACGTTGG 4492
Qy 737 CCCTGTCTGCTTCTGTAAACCTCCCTCCATC 767
Db 4493 CCATGTCTGCTTCTGTAAACCTCCCTCCATC 4523

RESULT 11

US-10-784-004-1017
; Sequence 1017, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784.004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1017
; LENGTH: 4752
; TYPE: DNA
; ORGANISM: human
US-10-784-004-1017

Query Match 96.4%; Score 743; DB 7; Length 4752;
Best Local Similarity 99.3%; Pred. No. 7.3e-189;
Matches 746; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 17 TGATGCCAAATGTGTTGGTGAACCTCGAGCTGGACACCACTCGAGGTGGACACCACTCAAGAGCCTGAG 76
Db 3773 TGATGCCAAATGTGTTGGTGAACCTCGAGCTGGACACCACTCGAGGTGGACACCACTCAAGAGCCTGAG 3832
Qy 77 CCAGCAGATCGAAGAACATCCGAGCCCGAGGGAGCGCGAAGAACCCCGCCCGCACTG 136
Db 3833 CCAGCAGATCGAAGAACATCCGAGCCCGAGGGCAGCGCAAGAACCCCGCCCGCACTG 3892
Qy 137 CCGTGACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3893 CCGTGACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3952
Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGAGAGCTGTGAGAGCTG 256
Db 3953 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGAGAGCTGTGAGAGCTG 4012
Qy 257 CGTGTACCCCACTCAGCCCACTGAGTGGCCCAAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 4013 CGTGTACCCCACTCAGCCCACTGAGTGGCCCAAGAACTGGTACATCAGCAAGAACCCCAA 4072
Qy 317 GGAACAAGAGGATGTCTGGTTCGGCGAGAGATGACCGATGGATTCCAGTTGAGTATGG 376
Db 4073 GGAACAAGAGGATGTCTGGTTCGGCGAGAGATGACCGATGGATTCCAGTTGAGTATGG 4132
Qy 377 CGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTTCTGGCCCTGATGTC 436
Db 4133 CGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTTCTGGCCCTGATGTC 4192
Qy 437 CACCGAGGCTCCCGAAGAACATCACTTCACTCAAGAAAGAGCGCTTCACTACATGAGACCA 496
Db 4193 CACCGAGGCTCCCGAAGAACATCACTTCACTCAAGAAAGAGCGCTTCACTACATGAGACCA 4252
Qy 497 GCAGACTGGCAACTCCAGAACATCACTTCACTCAAGAAAGAGCGCTTCACTACATGAGACCA 556
Db 4253 GCAGACTGGCAACTCCAGAACATCACTTCACTCAAGAAAGAGCGCTTCACTACATGAGACCA 4312
Qy 557 CGCGAGGGCAACAGCGCTTCACTACAGGCTTCACTTCACTGATGGCTGCAGAGTCAAC 616
Db 4313 CGCGAGGGCAACAGCGCTTCACTACAGGCTTCACTTCACTGATGGCTGCAGAGTCAAC 4372

Qy 617 CGGAGCCTGGGGCAAGACAGTGAATTAACAACCAACCAAGACCTCCCGCCTGGCCAT 676
Db 4373 CGGAGCCTGGGGCAAGACAGTGAATTAACAACCAACCAAGACCTCCCGCCTGGCCAT 4432
Qy 677 CATGATGTGGCCCTTGGAGCTGTGGTGGCCCAAGACCAAGAAATTCGGCTTCGACGTTGG 736
Db 4433 CATGATGTGGCCCTTGGAGCTGTGGTGGCCCAAGACCAAGAAATTCGGCTTCGACGTTGG 4492
Qy 737 CCCTGTCTGCTTCTGTAAACCTCCCTCCATC 767
Db 4493 CCATGTCTGCTTCTGTAAACCTCCCTCCATC 4523

RESULT 12

US-10-784-004-1195
; Sequence 1195, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784.004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1195
; LENGTH: 4752
; TYPE: DNA
; ORGANISM: human
US-10-784-004-1195

Query Match 96.4%; Score 743; DB 7; Length 4752;
Best Local Similarity 99.3%; Pred. No. 7.3e-189;
Matches 746; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 17 TGATGCCAAATGTGTTGGTGAACCTCGAGCTGGACACCACTCGAGGTGGACACCACTCAAGAGCCTGAG 76
Db 3773 TGATGCCAAATGTGTTGGTGAACCTCGAGCTGGACACCACTCGAGGTGGACACCACTCAAGAGCCTGAG 3832
Qy 77 CCAGCAGATCGAAGAACATCCGAGCCCGAGGGAGCGCGAAGAACCCCGCCCGCACTG 136
Db 3833 CCAGCAGATCGAAGAACATCCGAGCCCGAGGGCAGCGCAAGAACCCCGCCCGCACTG 3892
Qy 137 CCGTGACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 196
Db 3893 CCGTGACTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTGACCCCAA 3952
Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGAGAGCTGTGAGAGCTG 256
Db 3953 CCAAGGCTGCAACCTGGATGCCATCAAAAGTCTTCTGCAACATGAGAGCTGTGAGAGCTG 4012
Qy 257 CGTGTACCCCACTCAGCCCACTGAGTGGCCCAAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 4013 CGTGTACCCCACTCAGCCCACTGAGTGGCCCAAGAACTGGTACATCAGCAAGAACCCCAA 4072
Qy 317 GGAACAAGAGGATGTCTGGTTCGGCGAGAGATGACCGATGGATTCCAGTTGAGTATGG 376
Db 4073 GGAACAAGAGGATGTCTGGTTCGGCGAGAGATGACCGATGGATTCCAGTTGAGTATGG 4132
Qy 377 CGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTTCTGGCCCTGATGTC 436
Db 4133 CGGCGAGGGCTCCGACCTCGCGATGTGGCCATCCAGCTGACCTTCTTCTGGCCCTGATGTC 4192
Qy 437 CACCGAGGCTCCCGAAGAACATCACTTCACTCAAGAAAGAGCGCTTCACTACATGAGACCA 496
Db 4193 CACCGAGGCTCCCGAAGAACATCACTTCACTCAAGAAAGAGCGCTTCACTACATGAGACCA 4252
Qy 497 GCAGACTGGCAACTCCAGAACATCACTTCACTCAAGAAAGAGCGCTTCACTACATGAGACCA 556
Db 4253 GCAGACTGGCAACTCCAGAACATCACTTCACTCAAGAAAGAGCGCTTCACTACATGAGACCA 4312
Qy 557 CGCGAGGGCAACAGCGCTTCACTACAGGCTTCACTTCACTGATGGCTGCAGAGTCAAC 616

Db 4313 CGCCGAGGCAACAGCCGCTTACCTACAGCGTCACTGTCGATGGCTGCACGAGTCACAC 4372
Qy 617 CGGAGCTGGGGCAAGACAGTATTGAATCAAAACCCACCAAGTCTCTCCGCGCTGCCCAT 676
Db 4373 CGGAGCTGGGGCAAGACAGTATTGAATCAAAACCCACCAAGACCTCCCGCTTGGCAT 4432
Qy 677 CATCGATGTGGCCCTTGGACGTTGGTGCCCGCAGACACAGGAATTCGGCTTCGACGTTGG 736
Db 4433 CATCGATGTGGCCCTTGGACGTTGGTGCCCGCAGACACAGGAATTCGGCTTCGACGTTGG 4492
Qy 737 CCTGCTGCTTCTGTAACCTCCCTCCATC 767
Db 4493 CCATGCTGCTTCTGTAACCTCCCTCCATC 4523

RESULT 13

US-11-108-172-991
; Sequence 991, Application US/11108172
; Publication No. US20050260177A1
; GENERAL INFORMATION:
; APPLICANT: Xu Jiangchun
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Benson, Darin R.
; APPLICANT: Meagher, Madeleine Joy
; APPLICANT: Stolk, John A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Jiang, Yuqiu
; APPLICANT: Smith, Carole L.
; APPLICANT: King, Gordon E.
; APPLICANT: Wang, Aijun
; APPLICANT: Clapper, Jonathan D.
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick Thomas S.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF COLON CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121-471C15
; CURRENT APPLICATION NUMBER: US/11/108,172
; PRIOR FILING DATE: 2005-04-15
; PRIOR APPLICATION NUMBER: US 10/025,380
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 09/922,217
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 09/833,263
; PRIOR FILING DATE: 2001-04-10
; PRIOR APPLICATION NUMBER: US 09/649,811
; PRIOR FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: US 09/609,448
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 09/575,251
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/519,444
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: US 09/504,629
; PRIOR FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: US 09/480,321
; PRIOR FILING DATE: 2000-01-10
; PRIOR APPLICATION NUMBER: US 09/476,296
; PRIOR FILING DATE: 1999-12-30
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1130
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 991
; LENGTH: 659
; TYPE: DNA
; ORGANISM: Homo sapiens
US-11-108-172-991

Query Match 84.9%; Score 654.2; DB 14; Length 659;
Best Local Similarity 99.5%; Pred. No. 3.le-165;
Matches 656; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 50 GGTGGACACCCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGGG 109
Db 1 GGTGGACACCCCTCAAGAGCCTGAGCCAGCAGATCGAGAACATCCGAGGCCAGAGGG 60
Qy 110 AGCCCGCAAGAACCCCGCGGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAA 169
Db 61 CAGCCCAAGAACCCCGCGGACCTGCGGTGACCTCAAGATGTGCCACTCTGACTGGAA 120
Qy 170 GAGTGAGAGTACTGATTTGACCCCAACCAAGGCTCAACCTCGATGCGCATCAAGTCTT 229
Db 121 GAGTGAGAGTACTGATTTGACCCCAACCAAGGCTCAACCTCGATGCGCATCAAGTCTT 180
Qy 230 CTGCAACATGGAGACTGTGTGAGACCTGCGGTGTATCCCCACTCAGCCCAAGTGTGGCCAGAA 289
Db 181 CTGCAACATGGAGACTGTGTGAGACCTGCGGTGTATCCCCACTCAGCCCAAGTGTGGCCAGAA 240
Qy 290 GAACTGATCATCAGCAAGAACCCCAAGGACAAAGAGCATGTCTGGTTCGGCCAGAGCAT 349
Db 241 GAACTGATCATCAGCAAGAACCCCAAGGACAAAGAGCATGTCTGGTTCGGCCAGAGCAT 300
Qy 350 GACCGATGATTCACGATTCGAGTATGCGCGCCAGAGGCTCCGACCCCTGCGATGTGGCCAT 409
Db 301 GACCGATGATTCACGATTCGAGTATGCGCGCCAGAGGCTCCGACCCCTGCGATGTGGCCAT 360
Qy 410 CCAAGTGTGACCTTCTGCGCTGTATGTCCACCGAGGCTCCCGAGAACATCACTTACCATG 469
Db 361 CCAAGTGTGACCTTCTGCGCTGTATGTCCACCGAGGCTCCCGAGAACATCACTTACCATG 420
Qy 470 CAAGAACAGGTGGCTGTATGAGCAGACAGTGTGCGCAACCTCAAGAGGCCCTGCTCT 529
Db 421 CAAGAACAGGTGGCTGTATGAGCAGACAGTGTGCGCAACCTCAAGAGGCCCTGCTCT 480
Qy 530 CAGGGCTCCACGAGATCGAGATCGCGCCGAGGCAACAGCGCGTTCACCTACAGCGT 589
Db 481 CAGGGCTCCACGAGATCGAGATCGCGCCGAGGCAACAGCGCGTTCACCTACAGCGT 540
Qy 590 CACTGTGATGCTGTGACGAGTCAACCGAGGCTGCGGGCAAGACAGTGTGAATACAA 649
Db 541 CACTGTGATGCTGTGACGAGTCAACCGAGGCTGCGGGCAAGACAGTGTGAATACAA 600
Qy 650 AACCAACAGTCTCCCGCTGCGCATGATGTGGCCCTTTCGAGCGTTGGTGGCC 708
Db 601 AACCAACAGTCTCCCGCTGCGCATGATGTGGCCCTTTCGAGCGTTGGTGGCC 659

RESULT 14

US-10-784-004-81
; Sequence 81, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201-6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 81
; LENGTH: 5721
; TYPE: DNA
; ORGANISM: rat
US-10-784-004-81

Query Match 83.1%; Score 640.6; DB 7; Length 5721;
Best Local Similarity 90.8%; Pred. No. 2e-161;
Matches 682; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 17 TGATGCCAATGTGTTGTCGTCACCGTGTGACCTCGAGGTGGACCAACCCCTCAAGAGCCCTGAG 76
Db 3624 TGATGCCAATGTGTTGTCGTCACCGTGTGACCTCGAGGTGGACCAACCCCTCAAGAGCCCTGAG 3683
Qy 77 CCAGCAGATCGAGAACATATCCGAGGCCAGAGGAGCCGCAAGAACCCCGCCGACCTG 136

```
Db 3684 CCAGCAGATTCAGAACATCCGACGCTTGAGGGCAGCGCCAGAACCCCGCCGCAATG 3743
Qy 137 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTCAGCCCAA 196
Db 3744 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGCGGAGAGTACTGGATTCAGCCCTAA 3803
Qy 197 CCAAGGCTGCAACCTGGATGTCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGACTG 256
Db 3804 CCAAGGCTGCAACCTGGATGTCATCAAGTCTTCTGCAACATGGAGACAGGTCAAGACTG 3863
Qy 257 CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 3864 TGTGTTCCCACTCAGCCCTCTGTGCTCAGAGAACTGGTACATCAGCCCAACCCCAA 3923
Qy 317 GGAACAAGGCAATGTCGTGGTTCGGCGAGAGATGACCGATGATTCAGTTCGAGTATGG 376
Db 3924 GGAGAAAAGCATGTCGTGGTTCGGAGAGCATGACCGATGATTCAGTTCGAGTATGG 3983
Qy 377 CGGCCAGGGCTCCAGCCTGCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 436
Db 3984 AAGCGAAGGTTCCGATCTTCCGATGTGGCTATCCAGCTGACCTTCTGCGCCTGATGTC 4043
Qy 437 CACCGAGCCTCCAGAACATCACTACCTGCAAGAACAGCGTGGCTTACATGGACCA 496
Db 4044 CACCGAGCCTCCAGAACATCACTATCACTGCAAGAACAGCGTAGCCTTACATGGACCA 4103
Qy 497 GCAGACTGGCAACCTCAAGAGGCCCTGCTCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556
Db 4104 ACAGACTGGCAACCTCAAGAGGTCCTGCTCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 4163
Qy 557 CGCGAGGGCNAAGCGCTTCACTACAGGTCATCTGTCAGTGGCTGCGAGTCAAC 616
Db 4164 GGGCAGGGCAACAGTGCATTCACCTACAGCAGCGTTGTGGATGGCTGCAAGTCAAC 4223
Qy 617 CGGAGCTGGGGCAAGACAGTGAATTGAATACAAACCAAGTCCTCCCGCCTGCCAT 676
Db 4224 CGGAGCTGGGGCAAGACAGTGAATTGAATACAAACCAAGTCCTCCCGCCTGCCAT 4283
Qy 677 CATCGATGTGGCCCTTGGAGCGTTGGTGGCCCAAGACCAAGAAATTCGGCTTCGACGTTGG 736
Db 4284 CATCGATGTGGCTCCCTTGGACATTTGGTGGCCCAAGACCAAGAAATTCGGAAATGGACATTGG 4343
Qy 737 CCCTGTCTGCTTCGTGTAACCTCCCTCCATC 767
Db 4344 CCCTGTCTGCTTCGTGTAACCTCCCTCCATC 4374
```

RESULT 15

```
US-10-784-004-215
; Sequence 215, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10784.004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 215
; LENGTH: 5721
; TYPE: DNA
; ORGANISM: rat
US-10-784-004-215
```

```
Query Match 83.1%; Score 640.6; DB 7; Length 5721;
Best Local Similarity 90.8%; Pred. No. 2e-161;
Matches 682; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
Qy 17 TGATGCCAATGTGGTTCGTGACCGTGGACCTCGAGTGGACACCAACCCCTCAAGAGCCTGAG 76
Db 3624 TGATGCCAATGTGGTTCGTGACCGTGGACCTTGAGGTGGACACATACCTCAAGAGCCTGAG 3683
```

```
Search completed: April 27, 2006, 19:45:07
Job time : 532.452 secs
```

```
Qy 77 CCAGCAGATTCAGAACATCCGAGGCCAGAGGGAAGCGCAAGAACCCCGCCGCACTG 136
Db 3684 CCAGCAGATTCAGAACATCCGAGGCCCTGAGGGGAGCCGCAAGAACCCCGCCGCACTG 3743
Qy 137 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGTGGAGAGTACTGGATTCAGCCCAA 196
Db 3744 CCCTGACCTCAAGATGTGCCACTCTGACTGGAAGAGCGGAGAGTACTGGATTCAGCCCTAA 3803
Qy 197 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACTGGTGGAGACTG 256
Db 3804 CCAAGGCTGCAACCTGGATGCCATCAAGTCTTCTGCAACATGGAGACAGGTCAAGACTG 3863
Qy 257 CGTGTACCCCACTCAGCCAGTGTGGCCAGAGAACTGGTACATCAGCAAGAACCCCAA 316
Db 3864 TGTGTTCCCACTCAGCCCTCTGTGCTCAGAGAACTGGTACATCAGCCCAACCCCAA 3923
Qy 317 GGAACAAGGCAATGTCGTGGTTCGGCGAGAGCATGACCGATGATTCAGTTCGAGTATGG 376
Db 3924 GGAGAAAAGCATGTCGTGGTTCGGAGAGCATGACCGATGATTCAGTTCGAGTATGG 3983
Qy 377 CGGCCAGGGCTCCAGCCTGCGATGTGGCCATCCAGCTGACCTTCTGCGCCTGATGTC 436
Db 3984 AAGCGAAGGTTCCGATCTTCCGATGTGGCTATCCAGCTGACCTTCTGCGCCTGATGTC 4043
Qy 437 CACCGAGCCTCCAGAACATCACTACCTGCAAGAACAGCGTGGCTTACATGGACCA 496
Db 4044 CACCGAGCCTCCAGAACATCACTATCACTGCAAGAACAGCGTAGCCTTACATGGACCA 4103
Qy 497 GCAGACTGGCAACCTCAAGAGGCCCTGCTCTCTCAAGGGCTCCAAAGAGATCGAGATCCG 556
Db 4104 ACAGACTGGCAACCTCAAGAGTCCCTGCTCTCTCAAGGGCTCCAAAGAGATCGAGTCAAC 4163
Qy 557 CGCGAGGGCAACAGCGCTTCACTACAGCGTCACTGTGTCAGTGGCTGCGAGTCAAC 616
Db 4164 GGGCAGGGCAACAGTGCATTCACCTACAGCAGCGTTGTGGATGGCTGCAAGTCAAC 4223
Qy 617 CGGAGCCTGGGGCAAGACAGTGAATTGAATACAAACCAAGTCCTCCCGCCTGCCAT 676
Db 4224 CGGAGCCTGGGGCAAGACAGTGAATTGAATACAAACCAAGTCCTCCCGCCTGCCAT 4283
Qy 677 CATCGATGTGGCCCTTGGAGCGTTGGTGGCCCAAGACCAAGAAATTCGGCTTCGACGTTGG 736
Db 4284 CATCGATGTGGCTCCCTTGGACATTTGGTGGCCCAAGACCAAGAAATTCGGAAATGGACATTGG 4343
Qy 737 CCCTGTCTGCTTCGTGTAACCTCCCTCCATC 767
Db 4344 CCCTGTCTGCTTCGTGTAACCTCCCTCCATC 4374
```


THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:10:27 ; Search time 64.6272 Seconds
(without alignments)
1679.270 Million cell updates/sec

Title: US-10-677-877A-4
Perfect score: 1329
Sequence: 1 RSDANVVRDLEVDITLKS.....LDVGAPDQFEGDVGPCFL 247

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1329	100.0	247	9 AEA28571	Aea28571 Human col
2	1329	100.0	635	9 AEA28583	Aea28583 Human sol
3	1329	100.0	755	9 AEA28575	Aea28575 Human alk
4	1329	99.6	502	9 AEA28579	Aea28579 Human sol
5	1320	99.3	311	9 AEA28569	Aea28569 Human col
6	1320	99.3	566	9 AEA28577	Aea28577 Human sol
7	1320	99.3	699	9 AEA28581	Aea28581 Human sol
8	1320	99.3	819	9 AEA28573	Aea28573 Human alk
9	1320	99.3	1284	8 ADS98184	Ads98184 Protein f
10	1320	99.3	1464	4 AAB82454	Aab82454 Human pro
11	1320	99.3	1464	5 ABB90764	Abb90764 Human tum
12	1320	99.3	1464	5 ABP86610	Abp86610 Human pan
13	1320	99.3	1464	6 ABU54471	Abu54471 Human tum
14	1320	99.3	1464	6 ABR47417	Abr47417 Breast ca
15	1320	99.3	1464	6 ABR92064	Abr92064 Human cer
16	1320	99.3	1464	7 ADD14142	Add14142 Human src
17	1320	99.3	1464	7 ADP65246	Adp65246 Human alp
18	1320	99.3	1464	8 ADQ19470	Adq19470 Human sol
19	1320	99.3	1464	8 ADQ29653	Adq29653 Human col
20	1320	99.3	1464	9 ADV87000	Adv87000 Collagen
21	1320	99.3	1464	9 ADV70233	Adv70233 Tumox-ass
22	1320	99.3	1464	9 ADZ26573	Adz26573 Human typ
23	1320	99.3	1464	9 ADZ09874	Adz09874 Human bre
24	1320	99.3	1464	9 AEA04480	Aea04480 Human pro

25	1317	99.1	246	7 AAE38634	Aae38634 Human col
26	1317	99.1	250	2 AAW12845	Aaw12845 Pro-alpha
27	1317	99.1	1411	3 AAY56800	Aay56800 Human pre
28	1317	99.1	1461	5 ABG93947	Abg93947 Human pol
29	1317	99.1	1464	2 AAW68485	Aaw68485 Human rec
30	1317	99.1	1464	7 ADD45059	Add45059 Human pro
31	1317	99.1	1464	7 ADD45055	Add45055 Human pro
32	1317	99.1	1464	7 ADD45051	Add45051 Human pro
33	1317	99.1	1464	8 ADRI6800	Adri6800 Human col
34	1317	99.1	1464	8 ADRI6425	Adri6425 Human col
35	1317	99.1	1464	8 ADR99144	Adr99144 Collagen,
36	1317	99.1	1464	9 ADX58061	Adx58061 Amino aci
37	1315	98.9	258	7 ADD26744	Add26744 Human adi
38	1313	98.8	366	6 ABR41631	Abr41631 Human dit
39	1313	98.8	449	3 AAB43439	Aab43439 Human can
40	1313	98.8	465	7 ADE87049	Ade87049 Human pan
41	1313	98.8	886	7 ADE87052	Ade87052 Human pan
42	1313	98.8	1161	7 ADE87050	Ade87050 Human pan
43	1313	98.8	1464	4 AAU14136	Aau14136 Human nov
44	1313	98.8	1464	7 ADE87048	Ade87048 Human pan
45	1313	98.8	1536	7 ADE87051	Ade87051 Human pan

ALIGNMENTS

RESULT 1
AEA28571
ID AEA28571 standard; protein; 247 AA.

XX AC AEA28571;

XX DT 28-JUL-2005 (first entry)

XX DE Human collagen alpha(I) C-propeptide T2 construct protein.

XX KW protein secretion; fusion protein; protein activation; collagen I.

XX OS Homo sapiens.

XX OS Synthetic.

XX PN WO2005047850-A2.

XX PD 26-MAY-2005.

XX PF 04-OCT-2004; 2004WO-US032753.

XX PR 02-OCT-2003; 2003US-00677877.

XX PA (GENH-) GENHUNTER CORP.

XX PI Liang P;

XX DR WPI; 2005-386406/39.

XX DR N-PSDB; AEA28570.

XX PT Generating a secreted trimeric fusion protein by introducing into a cell a DNA construct comprising a promoter linked to a template encoding a signal peptide sequence followed by in-frame fusion to polypeptide to be trimerized.

XX PS Claim 14; SEQ ID NO 4; 47pp; English.

XX CC The invention relates to a novel method for generating a secreted trimeric fusion protein. The method comprises creating a DNA construct comprising a transcriptional promoter linked to a template encoding a signal peptide sequence followed, via an in-frame fusion, to the polypeptide to be trimerized, which in turn is joined in-frame to a polypeptide capable of self-trimerization which is heterologous to the first polypeptide to be trimerized, and introducing the construct into a eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of only the extracellular ligand-binding domain, may be used to intercept a ligand and thus overcome the overactivation of a receptor during disease

CC treatment. Trimeric receptor decoys theoretically should have a much
 CC higher affinity to their ligand than their dimeric counterparts. Such
 CC rationally designed soluble trimeric receptor analogs could significantly
 CC increase clinical benefits, as well as lower the amount or frequency of
 CC drug injections for each patient. The method of the invention may be
 CC useful for generating a secreted trimeric fusion protein to be used for
 CC more efficient neutralization of the biological activities of their
 CC naturally occurring trimeric ligands. The current sequence is that of the
 CC human collagen alpha(I) C-propeptide T2 construct protein of the
 CC invention. The construct includes a mutated BMP-1 recognition site but
 CC does not contain the glycine-repeat triple helical region.
 XX
 SQ Sequence 247 AA;

Query Match 100.0%; Score 1329; DB 9; Length 247;
 Best Local Similarity 100.0%; Pred. No. 2.5e-127;
 Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RSDANVVRDRDLEVDITLKSQQIENIRSPGSRKNPARTCDLKNCHSDMKSGEYWD 60
 DB 1 RSDANVVRDRDLEVDITLKSQQIENIRSPGSRKNPARTCDLKNCHSDMKSGEYWD 60
 QY 61 PNOGCNLDALKVFNCNMTGETCVYPTQPSVAQKNWYISKNPDKRHWFGESMTDGFQFE 120
 DB 61 PNOGCNLDALKVFNCNMTGETCVYPTQPSVAQKNWYISKNPDKRHWFGESMTDGFQFE 120
 QY 121 YGGQSGDPADVAIQLTFLRLMSTEAQNITYHCNLSVAYMDQQTGNLKKALLKGSNEIE 180
 DB 121 YGGQSGDPADVAIQLTFLRLMSTEAQNITYHCNLSVAYMDQQTGNLKKALLKGSNEIE 180
 QY 181 IRAEGNSRFTYSVTVDGCTSHTGAWKTVIEYTKSSRLPIIDVAPLDVGAPDQBFQFD 240
 DB 181 IRAEGNSRFTYSVTVDGCTSHTGAWKTVIEYTKSSRLPIIDVAPLDVGAPDQBFQFD 240

QY 241 VGPVCFL 247
 DB 241 VGPVCFL 247

RESULT 2

AEA28583
 ID AEA28583 standard; protein; 635 AA.

AC AEA28583;

XX 28-JUL-2005 (first entry)

DT Human soluble CD4-collagen alpha(I) C-propeptide T2 fusion protein.

DE protein secretion; fusion protein; protein activation; collagen I; CD4;
 KW mutein.

XX Homo sapiens.

OS Synthetic.

OS Chimeric.

XX WO2005047850-A2.

XX 26-MAY-2005.

XX 04-OCT-2004; 2004WO-US032753.

XX 02-OCT-2003; 2003US-00677877.

XX (GENH-) GENHUNTER CORP.

XX Liang P;

XX WPI; 2005-386406/39.

DR N-PSDB; AEA28582.

XX Generating a secreted trimeric fusion protein by introducing into a cell
 PT a DNA construct comprising a promoter linked to a template encoding a

PT signal peptide sequence followed by in-frame fusion to polypeptide to be
 PT trimerized.

XX Claim 16; SEQ ID NO 16; 47pp; English.

XX The invention relates to a novel method for generating a secreted
 CC trimeric fusion protein. The method comprises creating a DNA construct
 CC comprising a transcriptional promoter linked to a template encoding a
 CC signal peptide sequence followed, via an in-frame fusion, to the
 CC polypeptide to be trimerized, which in turn is joined in-frame to a
 CC polypeptide capable of self-crimerization which is heterologous to the
 CC first polypeptide to be trimerized, and introducing the construct into a
 CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
 CC only the extracellular ligand-binding domain, may be used to intercept a
 CC ligand and thus overcome the overactivation of a receptor during disease
 CC treatment. Trimeric receptor decoys theoretically should have a much
 CC higher affinity to their ligand than their dimeric counterparts. Such
 CC rationally designed soluble trimeric receptor analogs could significantly
 CC increase clinical benefits, as well as lower the amount or frequency of
 CC drug injections for each patient. The method of the invention may be
 CC useful for generating a secreted trimeric fusion protein to be used for
 CC more efficient neutralization of the biological activities of their
 CC naturally occurring trimeric ligands. The current sequence is that of the
 CC human soluble CD4-human collagen alpha(I) C-propeptide T2 fusion protein
 CC (sCD4-T2) of the invention. The collagen T2 construct includes a mutated
 CC BMP-1 recognition site but does not contain the glycine-repeat triple
 CC helical region.

XX Sequence 635 AA;

Query Match 100.0%; Score 1329; DB 9; Length 635;
 Best Local Similarity 100.0%; Pred. No. 1e-126;
 Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RSDANVVRDRDLEVDITLKSQQIENIRSPGSRKNPARTCDLKNCHSDMKSGEYWD 60

DB 389 RSDANVVRDRDLEVDITLKSQQIENIRSPGSRKNPARTCDLKNCHSDMKSGEYWD 448

QY 61 PNOGCNLDALKVFNCNMTGETCVYPTQPSVAQKNWYISKNPDKRHWFGESMTDGFQFE 120

DB 449 PNOGCNLDALKVFNCNMTGETCVYPTQPSVAQKNWYISKNPDKRHWFGESMTDGFQFE 508

QY 121 YGGQSGDPADVAIQLTFLRLMSTEAQNITYHCNLSVAYMDQQTGNLKKALLKGSNEIE 180

DB 509 YGGQSGDPADVAIQLTFLRLMSTEAQNITYHCNLSVAYMDQQTGNLKKALLKGSNEIE 568

QY 181 IRAEGNSRFTYSVTVDGCTSHTGAWKTVIEYTKSSRLPIIDVAPLDVGAPDQBFQFD 240

DB 569 IRAEGNSRFTYSVTVDGCTSHTGAWKTVIEYTKSSRLPIIDVAPLDVGAPDQBFQFD 628

QY 241 VGPVCFL 247

DB 629 VGPVCFL 635

RESULT 3

AEA28575

ID AEA28575 standard; protein; 755 AA.

XX AEA28575;

XX 28-JUL-2005 (first entry)

DE Human alkaline phosphatase-collagen alpha C-propeptide fusion protein 2.

XX protein secretion; fusion protein; protein activation; collagen I;

KW Alkaline phosphatase; enzyme; mutein.

XX Homo sapiens.

OS Synthetic.

OS Chimeric.

XX Key

Location/Qualifiers

FT Misc-difference 509 /note= "Encoded by GA"
 FT XX WO2005047850-A2.
 FN PD 26-MAY-2005.
 XX PF 04-OCT-2004; 2004WO-US032753.
 XX PR 02-OCT-2003; 2003US-00677877.
 XX PA (GENH-) GENHUNTER CORP.
 XX PI Liang P; ✓
 XX NPI; 2005-386406/39.
 DR N-PSDB; AEA28574.
 XX
 XX Generating a secreted trimeric fusion protein by introducing into a cell
 PT a DNA construct comprising a promoter linked to a template encoding a
 PT signal peptide sequence followed by in-frame fusion to polypeptide to be
 PT trimerized.
 XX
 XX Claim 17; SEQ ID NO 8; 47pp; English.
 XX
 XX The invention relates to a novel method for generating a secreted
 CC trimeric fusion protein. The method comprises creating a DNA construct
 CC comprising a transcriptional promoter linked to a template encoding a
 CC signal peptide sequence followed, via an in-frame fusion, to the
 CC polypeptide to be trimerized, which in turn is joined in-frame to a
 CC polypeptide capable of self-trimerization which is heterologous to the
 CC first polypeptide to be trimerized, and introducing the construct into a
 CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
 CC only the extracellular ligand-binding domain, may be used to intercept a
 CC ligand and thus overcome the overactivation of a receptor during disease
 CC treatment. Trimeric receptor decoys theoretically should have a much
 CC higher affinity to their ligand than their dimeric counterparts. Such
 CC rationally designed soluble trimeric receptor analogs could significantly
 CC increase clinical benefits, as well as lower the amount or frequency of
 CC drug injections for each patient. The method of the invention may be
 CC useful for generating a secreted trimeric fusion protein to be used for
 CC more efficient neutralization of the biological activities of their
 CC naturally occurring trimeric ligands. The current sequence is that of the
 CC human placental alkaline phosphatase (AP)-human collagen alpha(I) C-
 CC propeptide T2 fusion protein (AP-T2) of the invention. The collagen T2
 CC construct includes a mutated BMP-1 recognition site but does not contain
 CC the glycine-repeat triple helical region.
 XX
 SQ Sequence 755 AA;

 Query Match 100.0%; Score 1329; DB 9; Length 755;
 Best Local Similarity 100.0%; Pred. No. 1.4e-126;
 Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 1 RSDANVVRDRDLEVDVTTKLSLSQOIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWD 60
 DB 509 RSDANVVRDRDLEVDVTTKLSLSQOIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWD 568
 QY 61 PNOGCNLDALVFCNMTGETCVPTOPSAQKNWYISKNPDKRHHVWFGESMTDGFQFE 120
 DB 569 PNOGCNLDALVFCNMTGETCVPTOPSAQKNWYISKNPDKRHHVWFGESMTDGFQFE 628
 QY 121 YGGQGSDDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIE 180
 DB 629 YGGQGSDDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIE 688
 QY 181 IRAEGNSRFTYSVTVDCTSHTGAWGKTVEIYTKTKSSRLPIIDVAPLDVGPDPQERGF 240
 DB 689 IRAEGNSRFTYSVTVDCTSHTGAWGKTVEIYTKTKSSRLPIIDVAPLDVGPDPQERGF 748
 QY 241 VGPVCF 247
 DB 749 VGPVCF 755

RESULT 4
 AEA28579
 ID AEA28579 standard; protein; 502 AA.
 XX
 AC AEA28579;
 XX
 DT 28-JUL-2005 (first entry)
 XX
 DE Human soluble TNF-RII-collagen alpha(I) C-propeptide T2 fusion protein.
 XX
 KW protein secretion; fusion protein; protein activation; collagen I;
 KW Tumor necrosis factor; TNF-RII; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS Chimeric.
 XX
 PN WO2005047850-A2.
 XX
 PD 26-MAY-2005.
 XX
 PF 04-OCT-2004; 2004WO-US032753.
 XX
 PR 02-OCT-2003; 2003US-00677877.
 XX
 PA (GENH-) GENHUNTER CORP.
 XX
 PI Liang P;
 XX
 DR NPI; 2005-386406/39.
 DR N-PSDB; AEA28578.
 XX
 XX Generating a secreted trimeric fusion protein by introducing into a cell
 PT a DNA construct comprising a promoter linked to a template encoding a
 PT signal peptide sequence followed by in-frame fusion to polypeptide to be
 PT trimerized.
 XX
 XX Claim 15; SEQ ID NO 12; 47pp; English.
 XX
 XX The invention relates to a novel method for generating a secreted
 CC trimeric fusion protein. The method comprises creating a DNA construct
 CC comprising a transcriptional promoter linked to a template encoding a
 CC signal peptide sequence followed, via an in-frame fusion, to the
 CC polypeptide to be trimerized, which in turn is joined in-frame to a
 CC polypeptide capable of self-trimerization which is heterologous to the
 CC first polypeptide to be trimerized, and introducing the construct into a
 CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
 CC only the extracellular ligand-binding domain, may be used to intercept a
 CC ligand and thus overcome the overactivation of a receptor during disease
 CC treatment. Trimeric receptor decoys theoretically should have a much
 CC higher affinity to their ligand than their dimeric counterparts. Such
 CC rationally designed soluble trimeric receptor analogs could significantly
 CC increase clinical benefits, as well as lower the amount or frequency of
 CC drug injections for each patient. The method of the invention may be
 CC useful for generating a secreted trimeric fusion protein to be used for
 CC more efficient neutralization of the biological activities of their
 CC naturally occurring trimeric ligands. The current sequence is that of the
 CC human placental alkaline phosphatase (AP)-human collagen alpha(I) C-
 CC propeptide T2 fusion protein (AP-T2) of the invention. The collagen T2
 CC construct includes a mutated BMP-1 recognition site but does not contain
 CC the glycine-repeat triple helical region.
 XX
 SQ Sequence 502 AA;

 Query Match 99.8%; Score 1324; DB 9; Length 502;
 Best Local Similarity 100.0%; Pred. No. 2.4e-126;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 QY 2 SDANVVRDRDLEVDVTTKLSLSQOIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWD 61
 DB 257 SDANVVRDRDLEVDVTTKLSLSQOIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWD 316

QY 62 NOCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFSGESMTDGFQFEY 121
DB 317 NOCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFSGESMTDGFQFEY 376
QY 122 GQGSDDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 181
DB 377 GQGSDDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 436
QY 182 RAEGNSRFTYSVTVDGCTSHGTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDV 241
DB 437 RAEGNSRFTYSVTVDGCTSHGTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDV 496
QY 242 GPVCFL 247
DB 497 GPVCFL 502
RESULT 5
AEA28569
ID AEA28569 standard; protein; 311 AA.
XX AEA28569;
AC AEA28569;
XX 28-JUL-2005 (first entry)
XX Human collagen alpha(I) C-propeptide T0 construct protein.
DE Human collagen alpha(I) C-propeptide T0 construct protein.
KW protein secretion; fusion protein; protein activation; collagen I.
XX Homo sapiens.
OS Synthetic.
XX Key Location/Qualifiers
FT Region 1..40
FT /note= "Glycine repeat region"
XX WO2005047850-A2.
XX 26-MAY-2005.
XX 04-OCT-2004; 2004WO-US032753.
XX 02-OCT-2003; 2003US-00677877.
XX (GENH-) GENHUNTER CORP.
XX Liang P;
XX WPI; 2005-386406/39.
XX N-PSDB; AEA28568.
XX Generating a secreted trimeric fusion protein by introducing into a cell
PT a DNA construct comprising a promoter linked to a template encoding a
PT signal peptide sequence followed by in-frame fusion to polypeptide to be
PT trimerized.
XX Claim 11; SEQ ID NO 2; 47pp; English.
XX The invention relates to a novel method for generating a secreted
CC trimeric fusion protein. The method comprises creating a DNA construct
CC comprising a transcriptional promoter linked to a template encoding a
CC signal peptide sequence followed, via an in-frame fusion, to the
CC polypeptide to be trimerized, which in turn is joined in-frame to a
CC polypeptide capable of self-trimerization which is heterologous to the
CC first polypeptide to be trimerized, and introducing the construct into a
CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
CC only the extracellular ligand-binding domain, may be used to intercept a
CC ligand and thus overcome the overactivation of a receptor during disease
CC treatment. Trimeric receptor decoys theoretically should have a much
CC higher affinity to their ligand than their dimeric counterparts. Such
CC rationally designed soluble trimeric receptor analogs could significantly
CC increase clinical benefits, as well as lower the amount or frequency of

CC drug injections for each patient. The method of the invention may be
CC useful for generating a secreted trimeric fusion protein to be used for
CC more efficient neutralization of the biological activities of their
CC naturally occurring trimeric ligands. The current sequence is that of the
CC human collagen alpha(I) C-propeptide T0 construct protein of the
CC invention. The construct includes a partial glycine-repeat triple helical
CC region.
XX Sequence 311 AA;
SQ Query Match 99.3%; Score 1320; DB 9; Length 311;
Best Local Similarity 100.0%; Pred. No. 3e-126;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 DANVVRDRDLEVDTTLKSLSQOIEIRSPGGRKNPARTCRDLKMCCHSDWKSGEYWDPN 62
DB 67 DANVVRDRDLEVDTTLKSLSQOIEIRSPGGRKNPARTCRDLKMCCHSDWKSGEYWDPN 126
QY 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFSGESMTDGFQFEY 122
DB 127 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFSGESMTDGFQFEY 186
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
DB 187 GQGSDDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 246
QY 183 AEGNSRFTYSVTVDGCTSHGTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDV 242
DB 247 AEGNSRFTYSVTVDGCTSHGTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDV 306
QY 243 PVCFL 247
DB 307 PVCFL 311
RESULT 6
AEA28577
ID AEA28577 standard; protein; 566 AA.
XX AEA28577;
AC AEA28577;
XX 28-JUL-2005 (first entry)
XX Human soluble TNF-RII-collagen alpha(I) C-propeptide T0 fusion protein.
DE Human soluble TNF-RII-collagen alpha(I) C-propeptide T0 fusion protein.
KW protein secretion; fusion protein; protein activation; collagen I;
KW tumor necrosis factor; TNF-RII.
XX Homo sapiens.
OS Synthetic.
OS Chimeric.
XX Key Location/Qualifiers
FT Region 259..294
FT /label= Human collagen alpha(I) glycine repeat region
XX WO2005047850-A2.
XX 26-MAY-2005.
XX 04-OCT-2004; 2004WO-US032753.
XX 02-OCT-2003; 2003US-00677877.
XX (GENH-) GENHUNTER CORP.
XX Liang P;
XX WPI; 2005-386406/39.
XX N-PSDB; AEA28576.
XX Generating a secreted trimeric fusion protein by introducing into a cell
PT a DNA construct comprising a promoter linked to a template encoding a

PT signal peptide sequence followed by in-frame fusion to polypeptide to be
trimerized.

XX Claim 15; SEQ ID NO 10; 47pp; English.

PS The invention relates to a novel method for generating a secreted
XX trimeric fusion protein. The method comprises creating a DNA construct
CC comprising a transcriptional promoter linked to a template encoding a
CC signal peptide sequence followed, via an in-frame fusion, to the
CC polypeptide to be trimerized, which in turn is joined in-frame to a
CC polypeptide capable of self-trimerization which is heterologous to the
CC first polypeptide to be trimerized, and introducing the construct into a
CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
CC only the extracellular ligand-binding domain, may be used to intercept a
CC ligand and thus overcome the overactivation of a receptor during disease
CC treatment. Trimeric receptor decoys theoretically should have a much
CC higher affinity to their ligand than their dimeric counterparts. Such
CC rationally designed soluble trimeric receptor analogs could significantly
CC increase clinical benefits, as well as lower the amount or frequency of
CC drug injections for each patient. The method of the invention may be
CC useful for generating a secreted trimeric fusion protein to be used for
CC more efficient neutralization of the biological activities of their
CC naturally occurring trimeric ligands. The current sequence is that of the
CC human soluble TNP-R11-human collagen alpha(I) C-propeptide T0 fusion
CC protein (sTNP-R11-T0) of the invention. The collagen T0 construct
CC includes a partial glycine-repeat triple helical region.

XX SQ Sequence 566 AA;

Query Match 99.3%; Score 1320; DB 9; Length 566;
Best Local Similarity 100.0%; Pred. No. 7.4e-126; Indels 0; Gaps 0;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 DANVVRDRDLEVDITLKSLSQQIENIRSPGSRKNPARTCRDLKMHSDWKSGEYWDPN 62
Db 322 DANVVRDRDLEVDITLKSLSQQIENIRSPGSRKNPARTCRDLKMHSDWKSGEYWDPN 381
Qy 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWFGESMTDGFQPEYG 122
Db 382 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWFGESMTDGFQPEYG 441
Qy 123 QGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQOTGNLKKALLKGSNEIR 182
Db 442 QGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQOTGNLKKALLKGSNEIR 501
Qy 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 242
Db 502 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 561
Qy 243 PFCFL 247
Db 562 PFCFL 566

RESULT 7

AEA28581
ID AEA28581 standard; protein; 699 AA.

XX AC AEA28581;

XX DT 28-JUL-2005 (first entry)

XX DE Human soluble CD4-collagen alpha(I) C-propeptide T0 fusion protein.

XX KW protein secretion; fusion protein; protein activation; collagen I; CD4.

XX OS Homo sapiens.

OS Synthetic.

OS Chimeric.

XX FH Key Location/Qualifiers

FT Region 391..427

FT /label= Human collagen alpha(I) glycine repeat region

XX WO2005047850-A2.

XX PD 26-MAY-2005.

XX PF 04-OCT-2004; 2004WO-US032753.

XX PR 02-OCT-2003; 2003US-00677877.

XX PA (GENH-) GENHUNTER CORP.

XX PI Liang P;

XX WPI; 2005-386406/39.

DR N-PSDB; AEA28580.

XX Generating a secreted trimeric fusion protein by introducing into a cell
PT a DNA construct comprising a promoter linked to a template encoding a
PT signal peptide sequence followed by in-frame fusion to polypeptide to be
PT trimerized.

XX Claim 16; SEQ ID NO 14; 47pp; English.

XX The invention relates to a novel method for generating a secreted
CC trimeric fusion protein. The method comprises creating a DNA construct
CC comprising a transcriptional promoter linked to a template encoding a
CC signal peptide sequence followed, via an in-frame fusion, to the
CC polypeptide to be trimerized, which in turn is joined in-frame to a
CC polypeptide capable of self-trimerization which is heterologous to the
CC first polypeptide to be trimerized, and introducing the construct into a
CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
CC only the extracellular ligand-binding domain, may be used to intercept a
CC ligand and thus overcome the overactivation of a receptor during disease
CC treatment. Trimeric receptor decoys theoretically should have a much
CC higher affinity to their ligand than their dimeric counterparts. Such
CC rationally designed soluble trimeric receptor analogs could significantly
CC increase clinical benefits, as well as lower the amount or frequency of
CC drug injections for each patient. The method of the invention may be
CC useful for generating a secreted trimeric fusion protein to be used for
CC more efficient neutralization of the biological activities of their
CC naturally occurring trimeric ligands. The current sequence is that of the
CC human soluble CD4-human collagen alpha(I) C-propeptide T0 fusion protein
CC (sCD4-T0) of the invention. The collagen T0 construct includes a partial
CC glycine-repeat triple helical region.

XX SQ Sequence 699 AA;

Query Match 99.3%; Score 1320; DB 9; Length 699;
Best Local Similarity 100.0%; Pred. No. 1e-125; Indels 0; Gaps 0;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 DANVVRDRDLEVDITLKSLSQQIENIRSPGSRKNPARTCRDLKMHSDWKSGEYWDPN 62
Db 455 DANVVRDRDLEVDITLKSLSQQIENIRSPGSRKNPARTCRDLKMHSDWKSGEYWDPN 514
Qy 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWFGESMTDGFQPEYG 122
Db 515 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWFGESMTDGFQPEYG 574
Qy 123 QGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQOTGNLKKALLKGSNEIR 182
Db 575 QGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQOTGNLKKALLKGSNEIR 634
Qy 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 242
Db 635 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 694
Qy 243 PFCFL 247
Db 695 PFCFL 699

RESULT 8

AEA28573
 ID AEA28573 standard; protein; 819 AA.
 AC AEA28573;
 XX
 DT 28-JUL-2005 (first entry)
 XX
 DE Human alkaline phosphatase-collagen alpha(I) C-propeptide fusion protein.
 XX
 KW protein secretion; fusion protein; protein activation; collagen I;
 KW Alkaline phosphatase; enzyme.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Region 511..548
 FT /label= Human collagen alpha(I) glycine repeat region
 XX
 PN WO2005047850-A2.
 XX
 PD 26-MAY-2005.
 XX
 PF 04-OCT-2004; 2004WO-US032753.
 XX
 PR 02-OCT-2003; 2003US-00677877.
 XX
 PA (GENH-) GENHUNTER CORP.
 XX
 PI Liang P;
 XX
 DR WPI; 2005-386406/39.
 DR N-PSDB; AEA28572.
 XX
 XX
 PT Generating a secreted trimeric fusion protein by introducing into a cell
 PT a DNA construct comprising a promoter linked to a template encoding a
 PT signal peptide sequence followed by in-frame fusion to polypeptide to be
 PT trimerized.
 XX
 PS Claim 17; SEQ ID NO 6; 47pp; English.
 XX
 CC The invention relates to a novel method for generating a secreted
 CC trimeric fusion protein. The method comprises creating a DNA construct
 CC comprising a transcriptional promoter linked to a template encoding a
 CC signal peptide sequence followed, via an in-frame fusion, to the
 CC polypeptide to be trimerized, which in turn is joined in-frame to a
 CC polypeptide capable of self-trimerization which is heterologous to the
 CC first polypeptide to be trimerized, and introducing the construct into a
 CC eukaryotic cell. Receptor decoys, e.g. soluble receptors consisting of
 CC only the extracellular ligand-binding domain, may be used to intercept a
 CC ligand and thus overcome the overactivation of a receptor during disease
 CC treatment. Trimeric receptor decoys theoretically should have a much
 CC higher affinity to their ligand than their dimeric counterparts. Such
 CC rationally designed soluble trimeric receptor analogs could significantly
 CC increase clinical benefits, as well as lower the amount or frequency of
 CC drug injections for each patient. The method of the invention may be
 CC useful for generating a secreted trimeric fusion protein to be used for
 CC more efficient neutralization of the biological activities of their
 CC naturally occurring trimeric ligands. The current sequence is that of the
 CC human placental alkaline phosphatase (AP)-human collagen alpha(I) C-
 CC propeptide T0 fusion protein (AP-T0) of the invention. The collagen T0
 CC construct includes a partial glycine-repeat triple helical region.
 XX
 SQ Sequence 819 AA;
 Query Match 99.3%; Score 1320; DB 9; Length 819;
 Best Local Similarity 100.0%; Pred. No. 1.3e-125;
 Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 DANVVRDRDLEVTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYIDPN 62
 Db
 575 DANVVRDRDLEVTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYIDPN 634

63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKPNKDKRHVWFGESMTDGFQFEYG 122
 |||||
 635 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKPNKDKRHVWFGESMTDGFQFEYG 694
 |||||
 123 QGGSDDPADVAIQLTFLRLMSTEAQNITHCNKSVAYMDOQTGNLKKALLKGSNEIEIR 182
 |||||
 695 QGGSDDPADVAIQLTFLRLMSTEAQNITHCNKSVAYMDOQTGNLKKALLKGSNEIEIR 754
 |||||
 183 AEGNSRPTSVTVGCTSHTGAGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 242
 |||||
 755 AEGNSRPTSVTVGCTSHTGAGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 814
 |||||
 243 PFCFL 247
 |||||
 815 PFCFL 819
 |||||
 RESULT 9
 ADS98184
 ID ADS98184 standard; protein; 1284 AA.
 XX
 AC ADS98184;
 XX
 DT 30-DEC-2004 (first entry)
 XX
 DE Protein factor discovery related isolated human polypeptide, SEQ ID 448.
 XX
 KW antiinflammatory; cytostatic; antimicrobial; gene therapy; inflammation;
 KW leukaemia; nervous system disorder; infection.
 XX
 OS Homo sapiens.
 XX
 PN WO2004087874-A2.
 XX
 PD 14-OCT-2004.
 XX
 PF 24-MAR-2004; 2004WO-US009202.
 XX
 PR 28-MAR-2003; 2003US-0458824P.
 XX
 PA (NUVE-) NUVELO INC.
 PA (DRMA/) DRMANAC R T.
 XX
 PI Tang YT, Zhou P, Wang J, Wang ZW, Hu T;
 XX
 DR WPI; 2004-737686/72.
 DR N-PSDB; ADS97949.
 XX
 PT New polynucleotides encoding a polypeptide with biological activity,
 PT useful for treating inflammation, leukemias, nervous system disorders, or
 PT infections.
 XX
 PS Claim 20; SEQ ID NO 448; 253pp; English.
 XX
 CC The invention relates to a novel isolated polynucleotide comprising any
 CC of the 235 nucleotide sequences described in the specification. The
 CC invention further comprises: an isolated polynucleotide encoding a
 CC polypeptide with biological activity, where the polynucleotide hybridizes
 CC to one of the 235 novel polynucleotides under stringent hybridization
 CC conditions, or having greater than about 99% sequence identity with the
 CC novel polynucleotide; a vector comprising a novel polynucleotide; an
 CC expression vector comprising the novel polynucleotide; a host cell
 CC genetically engineered to comprise the novel polynucleotide, which can be
 CC operatively associated with a regulatory sequence that modulates
 CC expression of the polynucleotide in the host cell; an isolated
 CC polypeptide encoded by the novel polynucleotide, or a polynucleotide
 CC hybridizing under stringent conditions to the novel polynucleotide; a
 CC composition comprising the polypeptide and a carrier; an antibody
 CC directed against the polypeptide; a method for detecting the novel
 CC polynucleotide in a sample; a method for detecting the polypeptide in a
 CC sample; a method for identifying a compound that binds to the polypeptide
 CC ; a method for producing the polypeptide; an isolated polypeptide

comprising any of the 235 amino acid sequences described in the specification; and a collection of polynucleotides comprising of at least one of the polynucleotides cited above. The polypeptides and polynucleotides of the invention have antiinflammatory, cytostatic, and antimicrobial activities. The novel polynucleotide may be used to treat disorders by gene therapy. The polypeptides and polynucleotides are useful for treating inflammation, leukaemia, nervous system disorders, or infections. This sequence represents one of the 235 novel isolated polypeptides of the invention.

SQ Sequence 1284 AA;

Query Match 99.3%; Score 1320; DB 8; Length 1284;
Best Local Similarity 100.0%; Pred. No. 2.5e-125; Indels 0; Gaps 0;
Matches 245; Conservative 0; Mismatches 0;

QY 3 DANVVRDRDLEVDITLKSLSQIENIRSPGSRKNPARTCRDLKCHSDMKSGEYWDPN 62
DB 1040 DANVVRDRDLEVDITLKSLSQIENIRSPGSRKNPARTCRDLKCHSDMKSGEYWDPN 1099
QY 63 QGCNLDIAKVFPCNMEGTCTCYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 122
DB 1100 QGCNLDIAKVFPCNMEGTCTCYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 1159
QY 123 QGGSDDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
DB 1160 QGGSDDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1219
QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQGFQFDVG 242
DB 1220 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQGFQFDVG 1279
QY 243 PVCFL 247
DB 1280 PVCFL 1284

RESULT 10
AAB82454
ID AAB82454 standard; protein; 1464 AA.

XX AC AAB82454;

XX AC AAB82454;

XX DT 22-AUG-2001 (first entry)

XX DE Human pro-alpha-1 chain of type I procollagen.

XX KW COL1A1 gene; collagen; procollagen; human.

XX OS Homo sapiens.

XX FT Key Location/Qualifiers

XX FT Peptide 1..22

XX FT Protein /label= Signal_peptide

XX FT Protein 23..1464

XX FT Protein /label= Mature_protein

XX PN WO200144455-A2.

XX XX 21-JUN-2001.

XX XX 12-DEC-2000; 2000WO-GB004741.

XX XX 15-DEC-1999; 99GB-00029487.

XX XX (ASTR) ASTRAZENECA AB.

XX XX (ASTR) ASTRAZENECA UK LTD.

XX XX Berl R;

XX XX WPI; 2001-398145/42.

XX XX N-PSDB; AAF90491.

PT Novel antisense DNA oligonucleotide useful for inhibiting the expression of wild type COL1A1 gene, for treating, reducing the risk of, and preventing collagen disorders.

XX PS Disclosure; Page 21-26; 30pp; English.

XX CC The present sequence is that of the pro-alpha-1 chain of human type I procollagen. The present invention relates to antisense oligonucleotides (ASOs) and their use in inhibiting expression of type I procollagen. The ASOs comprise 18-25 nucleotides and are complementary to a specific region within the type I collagen pro-alpha-1 chain gene (see AAF90491), especially those given in AAF90492-503. They are capable of inhibiting the expression of the pro-alpha-1 chain in a cell that expresses it. The ASOs are used in a claimed method of treating, or reducing a risk of, a collagen disorder. Such disorders may include those caused by overproduction of collagen fibres, such as liver cirrhosis, kidney, liver and heart fibrosis, scleroderma, hypertrophic scars and keloids

XX SQ Sequence 1464 AA;

Query Match 99.3%; Score 1320; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.1e-125;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLEVDITLKSLSQIENIRSPGSRKNPARTCRDLKCHSDMKSGEYWDPN 62
DB 1220 DANVVRDRDLEVDITLKSLSQIENIRSPGSRKNPARTCRDLKCHSDMKSGEYWDPN 1279
QY 63 QGCNLDIAKVFPCNMEGTCTCYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 122
DB 1280 QGCNLDIAKVFPCNMEGTCTCYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 1339
QY 123 QGGSDDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
DB 1340 QGGSDDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1399
QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQGFQFDVG 242
DB 1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQGFQFDVG 1459
QY 243 PVCFL 247
DB 1460 PVCFL 1464

RESULT 11
ABB90764

ID ABB90764 standard; protein; 1464 AA.

XX AC ABB90764;

XX AC ABB90764;

XX DT 30-MAY-2002 (first entry)

XX DE Human Tumour Endothelial Marker polypeptide SEQ ID NO 261.

XX KW Human; mouse; rat; TEM; tumour endothelial marker; NEM; PEM; cytostatic; normal endothelial marker; pan-endothelial marker; immunostimulant; antiangiogenic; tumour; neoangiogenesis; vascularised tumour; polycystic kidney disease; diabetes; retinopathy; rheumatoid arthritis; psoriasis.

XX XX Homo sapiens.

XX XX WO200210217-A2.

XX XX 07-FEB-2002.

XX XX 01-AUG-2001; 2001WO-US024031.

XX XX 02-AUG-2000; 2000US-0222599P.

XX XX 11-AUG-2000; 2000US-0224360P.

XX XX 11-APR-2001; 2001US-0282850P.

XX XX

PA (UYJO) UNIV JOHNS HOPKINS.
XX St Croix B, Kinzler KW, Vogelstein B;
XX WPI; 2002-291856/33.
DR N-PSDB; ABL92119.
XX An isolated molecule comprising an antibody variable region which
PT specifically binds to an extracellular domain of a tumor endothelial
PT marker (TEM) protein, useful for inhibiting tumor growth.
XX Claim 54; Page 265-268; 331pp; English.
XX The invention relates to an isolated molecule comprising an antibody
CC variable region which specifically binds to an extracellular domain of a
CC tumour endothelial marker (TEM) protein selected from ABB90732, ABB90740,
CC ABB90749, ABB90750 and ABB90769. The antibodies which bind to TEM
CC proteins have cytostatic, immunostimulant and antiangiogenic activity.
CC They are useful for inhibiting tumour growth, neoangiogenesis in subjects
CC bearing a vascularised tumour, polycystic kidney disease, diabetic
CC retinopathy, rheumatoid arthritis and psoriasis. Human, mouse and rat TEM
CC genes and the encoded proteins (ABL92075-ABL92141 and ABB90721-ABB90789)
CC are disclosed, as are marker oligonucleotide sequences: tumour
CC endothelial markers (TEM) ABL91996-ABL92041 and ABL92191; normal
CC endothelial markers (NEM) ABL92042-ABL92074; and pan-endothelial markers
CC (PEM) ABL91903-ABL91995
XX Sequence 1464 AA;
SQ
Query Match 99.3%; Score 1320; DB 5; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.1e-125;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 DANVVRDRDLEVDVTTLKSLSSQIENIRSPGSRKNPARTCRDLKCHSDWKSGEYIDPN 62
Db 1220 DANVVRDRDLEVDVTTLKSLSSQIENIRSPGSRKNPARTCRDLKCHSDWKSGEYIDPN 1279
QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 1280 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 1339
QY 123 QGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
Db 1340 QGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1399
QY 243 PVCFL 247
Db 1460 PVCFL 1464
RESULT 12
ABP68610
ID ABP68610 standard; protein; 1464 AA.
XX
AC ABP68610;
XX
DT 14-JAN-2003 (first entry)
XX
DE Human pancreatic cancer expressed protein SEQ ID NO 159.
XX
KW Human; pancreas; cancer; gene therapy; vaccine; immunostimulant;
KW cytostatic; tumour.
XX
OS Homo sapiens.
XX
FN WO200260317-A2.
XX
PD 08-AUG-2002.
XX

PF 30-JAN-2002; 2002WO-US002781.
XX
PR 30-JAN-2001; 2001US-0265305P.
PR 31-JAN-2001; 2001US-0265882P.
PR 09-FEB-2001; 2001US-0267568P.
PR 21-MAR-2001; 2001US-0278651P.
PR 28-APR-2001; 2001US-0287112P.
PR 16-MAY-2001; 2001US-0291631P.
PR 12-JUL-2001; 2001US-0305484P.
PR 20-AUG-2001; 2001US-0313999P.
PR 27-NOV-2001; 2001US-0333626P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Benson DR, Kalos MD, Lodes MJ, Persing DH, Hepler WT, Jiang Y;
XX WPI; 2002-627435/67.
DR N-PSDB; ABV94755.
XX
XX New isolated polynucleotide and pancreatic tumor polypeptides, useful for
PT diagnosing, preventing and/or treating cancer, particularly pancreatic
PT cancer.
XX Claim 2; SEQ ID NO 159; 300pp + Sequence Listing; English.
PS The invention relates to an isolated polynucleotide (I) comprising: (a)
XX a group of over 4000 nucleotide sequences (ABV94628-ABV99145); (b)
CC complements of (a); (c) sequences consisting of at least 20 contiguous
CC residues of (a); (d) sequences that hybridize to (a), under moderately
CC stringent conditions; (e) sequences having at least 75% or 90% identity
CC to (a); or (f) degenerate variants of (a). Polypeptides (ABP68596-
CC ABP68637) encoded by (I) and oligonucleotide can be used to detect cancer
CC in a patient and compositions comprising polypeptides, polynucleotides,
CC antibodies, fusion proteins, T cell populations and antigen presenting
CC cells expressing the polypeptide are useful in treating pancreatic cancer
CC and stimulating an immune response. The polynucleotides can be used as
CC probes or primers for nucleic acid hybridisation, in the design and
CC preparation of ribozyme molecules for inhibiting expression of the tumour
CC polypeptides and proteins in the tumour cells, in vaccines and for gene
CC therapy. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 1464 AA;
Query Match 99.3%; Score 1320; DB 5; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.1e-125;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 DANVVRDRDLEVDVTTLKSLSSQIENIRSPGSRKNPARTCRDLKCHSDWKSGEYIDPN 62
Db 1220 DANVVRDRDLEVDVTTLKSLSSQIENIRSPGSRKNPARTCRDLKCHSDWKSGEYIDPN 1279
QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 1280 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 1339
QY 123 QGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
Db 1340 QGSDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1399
QY 183 AEGNSRFTYSVTVDGCTSHTGAWGKTVEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGAWGKTVEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 1459
QY 243 PVCFL 247
Db 1460 PVCFL 1464
RESULT 13
ABU54471
ID ABU54471 standard; protein; 1464 AA.

XX AC ABU54471;
XX DT 12-MAR-2003 (first entry)
XX DE Human tumour endothelial marker TEM 38.
XX KW Human; endothelial cell; EC; tumour endothelial cell; TEM; NEM;
KW Tumour endothelial marker; normal endothelial marker; PEM;
KW pan-endothelial marker; polycystic kidney disease; psoriasis;
KW diabetic retinopathy; rheumatoid arthritis; tumour angiogenesis;
KW neovascularization; immune response; cytostatic; antiangiogenic;
KW ophthalmological; antirheumatic; antiarthritic; antipsoriatic.
XX OS Homo sapiens.
XX PN WO200283874-A2.
XX PD 24-OCT-2002.
XX PX 10-APR-2002; 2002WO-US008253.
XX PR 11-APR-2001; 2001US-0282850P.
XX PR 06-FEB-2002; 2002US-0354262P.
XX PX (UJJO) UNIV JOHNS HOPKINS.
XX PA Carson-Walter E, St Croix B, Kinzler KW, Vogelstein B;
XX PI 11-APR-2001; 2001US-0282850P.
XX DR WPI; 2003-093016/08.
XX DR N-PSDB; ABX72044.
XX PT New purified human transmembrane protein, designated as tumor endothelial
PT marker (TEM) 3, useful for detecting, diagnosing or treating tumors,
PT polycystic kidney disease, diabetic retinopathy, rheumatoid arthritis or
PT psoriasis.
XX PS Disclosure; Page 295-297; 374pp; English.
XX CC The present invention relates to a novel method for the isolation of
CC endothelial cells (ECs), and the identification of genes expressed in
CC normal and tumour ECs. Tumour endothelial marker (TEM), normal
CC endothelial marker (NEM), and pan-endothelial marker (PEM) genes are
CC identified in human ECs. The human EC marker proteins and the
CC polynucleotide sequences encoding them are useful for detecting,
CC diagnosing or treating tumours as well as polycystic kidney disease,
CC diabetic retinopathy, rheumatoid arthritis, and psoriasis. They are also
CC useful for inhibiting neoangiogenesis or tumour angiogenesis, for
CC inducing an immune response to tumour endothelial cells in a patient, or
CC for identifying candidate drugs for treating tumours. The present
CC sequence represents a human TEM or NEM protein of the invention
XX SQ Sequence 1464 AA;
Query Match 99.3%; Score 1320; DB 6; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.1e-125;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
DB 1220 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1279
QY 63 QGCNLDIAKVFNCNETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 122
DB 1280 QGCNLDIAKVFNCNETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 1339
QY 123 QGSDPADVAQLTFLRLMSTEAQNITYHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 182
DB 1340 QGSDPADVAQLTFLRLMSTEAQNITYHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 1399
QY 183 AEGNSRFTYSVTVDGCTSHTGAWKTVIEYKTKSRPLIIDVAPLDVGPADQEGFDVG 242
DB 1400 AEGNSRFTYSVTVDGCTSHTGAWKTVIEYKTKSRPLIIDVAPLDVGPADQEGFDVG 1459

QY 243 PVCFL 247
DB 1460 PVCFL 1464
RESULT 14
ABR47417
ID ABR47417 standard; protein; 1464 AA.
XX AC ABR47417;
XX DT 12-JUN-2003 (first entry)
XX DE Breast cancer associated protein sequence SEQ ID NO:65.
XX KW Human; breast cancer; cytostatic; gene therapy.
XX OS Homo sapiens.
XX PN WO2003004989-A2.
XX PD 16-JAN-2003.
XX PF 21-JUN-2002; 2002WO-US019669.
XX PX 21-JUN-2001; 2001US-0299887P.
XX PR 27-JUN-2001; 2001US-0301572P.
XX PR 18-JUL-2001; 2001US-0306501P.
XX PR 25-SEP-2001; 2001US-0325002P.
XX PR 05-MAR-2002; 2002US-0362885P.
XX PR 14-MAY-2002; 2002US-0380391P.
XX PA (MILL-) MILLENIUM PHARM INC.
XX PI Lillie J, Gannavarapu M, Glatt K, Hoersch S, Kamatkar S;
PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
PI Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
XX WPI; 2003-210381/20.
XX DR N-PSDB; ACC50108.
XX CC Breast cancer diagnosis or treatment by comparing the level of expression
PT of a marker in a patient sample with that in the control non-breast
PT cancer sample.
XX PS Claim 1; SEQ ID NO 65; 128pp; English.
XX CC The present invention describes a method for assessing whether a patient
CC is afflicted with breast cancer. The method comprises comparing the level
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC ABR47386 to ABR47632) in a patient sample and the normal level of
CC expression of the marker in a control non-breast cancer sample, where a
CC significant increase in the level of expression of the marker in the
CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences from
CC the present invention have cytostatic activities and can be used in gene
CC therapy. The method is useful for diagnosing and treating breast cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1464 AA;
Query Match 99.3%; Score 1320; DB 6; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.1e-125;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
DB 1220 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1279
QY 63 QGCNLDIAKVFNCNETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 122

|||||
1280 QGCNLDKAIKVCNMTGCTCVYPTQPSVAQKNWYISKNPDKRHRVWFGESMTDGFQFEYG 1339
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIR 182
Db 1340 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIR 1399
QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 1459
QY 243 PVCFL 247
Db 1460 PVCFL 1464

RESULT 15
ABR92064
ID ABR92064 standard; protein; 1464 AA.
XX ABR92064;
AC ABR92064;
DT 10-SEP-2003 (first entry)
XX Human cervical cancer cell marker protein SEQ ID NO:36.
DE Human; cervical cancer; cervical cancer marker; cancer therapy;
KW detection; gene therapy; vaccine.
XX Homo sapiens.
XX WO2002101075-A2.
XX 19-DEC-2002.
XX 12-JUN-2002; 2002WO-US018638.
XX 13-JUN-2001; 2001US-0298155P.
PR 13-JUN-2001; 2001US-0298159P.
PR 14-NOV-2001; 2001US-0335936P.
XX (MILL-) MILLENNIUM PHARM INC.
XX Schlegel R, Chen Y, Zhao X, Monahan JE, Kamatkar S;
PI Gannavarapu M, Glatk K, Hoersch S;
XX WPI; 2003-156967/15.
DR N-PSDB; ACF12845.

XX New isolated nucleic acid molecule useful for detecting, characterizing,
PT preventing and treating human cervical cancers, in various prognostic and
PT diagnostic assays, in pharmacogenomics and in monitoring clinical trials.
XX Claim 4; Page 164-167; 386pp; English.
XX ACF12828 to ACF12947 encode the human cervical cancer marker proteins (I)
CC given in ABR92047 to ABR92164. A higher level of expression of (I) than
CC normal indicates the presence of cervical cancer. Also described: (I) a
CC vector (II) containing (I); (2) a host cell (III) containing (I); and (3)
CC assessing (M1) whether a patient is afflicted with cervical cancer,
CC comprising comparing the level of expression of a marker in a patient's
CC sample, and the normal level of expression of the marker in a control non
CC -cervical cancer sample, where a significant increase in the level of
CC expression of the marker in the patient's sample relative to that in the
CC control sample is an indication that the patient is afflicted with
CC cervical cancer. (I) has cytostatic activity, and can be used in gene
CC therapy and in vaccines. (I) is useful in detecting, characterizing,
CC preventing and treating human cervical cancers. (I) may also be used in
CC various prognostic and diagnostic assays, pharmacogenomics and in
CC monitoring clinical trials
XX Sequence 1464 AA;

Query Match 99.3%; Score 1320; DB 6; Length 1464;
Best Local Similarity 100.0%; Pred. No. 3.1e-125;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 DANVVRDRDLEVDVTTLKSLSQOIENIRSPGSRKKNPARTCRDLKMSHSDWKSGEYWDPN 62
Db 1220 DANVVRDRDLEVDVTTLKSLSQOIENIRSPGSRKKNPARTCRDLKMSHSDWKSGEYWDPN 1279
QY 63 QGCNLDKAIKVCNMTGCTCVYPTQPSVAQKNWYISKNPDKRHRVWFGESMTDGFQFEYG 122
Db 1280 QGCNLDKAIKVCNMTGCTCVYPTQPSVAQKNWYISKNPDKRHRVWFGESMTDGFQFEYG 1339
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIR 182
Db 1340 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIR 1399
QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 1459
QY 243 PVCFL 247
Db 1460 PVCFL 1464

Search completed: April 26, 2006, 16:13:37
Job time : 64.6272 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:14:30 ; Search time 22.5753 Seconds
(without alignments)
1052.725 Million cell updates/sec

Title: US-10-677-877a-4
Perfect score: 1329
Sequence: 1 RSDANVVRDRDLEVDITLKS.....LDVGAPDQERGFDPVPCFL 247

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1310	98.6	1464	1 CGHUIS	collagen alpha 1(I)
2	1225	92.2	473	2 I50629	collagen - chicken
3	1218	91.6	1453	2 S21626	collagen alpha 1(I)
4	1006.5	75.7	288	1 CGCH6C	collagen alpha 1(I)
5	982.5	73.9	1487	1 CGHUC6	collagen alpha 1(I)
6	979.5	73.7	1418	2 T45467	collagen alpha 1(I)
7	967.5	72.8	1419	2 A41182	collagen alpha 1(I)
8	967.5	72.8	1487	2 B41182	collagen alpha 1(I)
9	946.5	71.2	1486	1 B40333	collagen alpha 1(I)
10	937.5	70.5	464	2 S59513	collagen II A1 pro
11	929	69.9	1492	2 A40333	collagen alpha 1(I)
12	883	66.4	323	2 A61396	collagen alpha 1(I)
13	883	66.4	964	1 CGCH2S	collagen alpha 2(I)
14	882	66.4	365	2 S10847	collagen alpha 2(I)
15	859.5	64.7	1466	1 CGHU7L	collagen alpha 1(I)
16	854.5	64.3	1464	2 S59856	collagen alpha 1(I)
17	853.5	64.2	636	2 S41087	collagen alpha 1(I)
18	853	64.2	1366	1 CGHU2S	collagen alpha 2(I)
19	851.5	64.1	615	2 A05269	collagen alpha 1(I)
20	796.5	59.9	1373	1 A43291	collagen alpha 2(I)
21	755	56.8	1496	1 CGHU2V	collagen alpha 2(V)
22	744	56.0	1497	2 I49607	procollagen type V
23	714.5	53.8	673	1 CGB08C	collagen alpha 1(I)
24	663	49.9	143	2 S07234	collagen alpha 1(I)
25	508.5	38.3	1806	1 CGHU1E	collagen alpha 1(X)
26	507.5	38.2	1838	1 CGHU1V	collagen alpha 1(V)
27	505.5	38.0	1843	2 S18803	collagen alpha 1(V)
28	501.5	37.7	888	2 S28791	collagen alpha 1(X)
29	500	37.6	730	2 A36226	collagen alpha 1 c

30	496	37.3	632	2	S42731	collagen alpha 1 c
31	495	37.2	1414	1	S23809	collagen alpha 2(I)
32	488.5	36.8	482	2	B31795	collagen alpha 1(X)
33	471	35.4	3198	2	A43426	collagen alpha 2 f
34	397.5	29.9	1051	2	A35763	collagen alpha 2 c
35	353	26.6	91	2	S71545	alpha 2(I) procoll
36	352.5	26.5	547	2	A36046	collagen alpha cha
37	311	23.4	1546	1	CGHU2E	collagen alpha 2(X)
38	291	21.9	193	2	S07133	collagen alpha 1(I)
39	262	19.7	69	2	I46677	collagen alpha 2(I)
40	203	15.3	39	2	I65269	collagen alpha 1(I)
41	136	10.2	31	2	I50176	alpha collagen typ
42	129	9.7	37	2	I65271	collagen alpha 1(I)
43	104	7.8	21	2	I65270	collagen alpha 1(I)
44	90	6.8	293	2	T26934	hypothetical prote
45	89	6.7	884	2	T29637	hypothetical prote

ALIGNMENTS

RESULT 1

CGHUIS

collagen alpha 1(I) chain precursor - human

N;Alternate names: procollagen alpha 1(I) chain

C;Species: Homo sapiens (man)

C;Date: 12-Aug-1981 #sequence.revision 04-Oct-1996 #text change 31-Dec-2004

C;Accession: I60114; S01143; A93335; I55254; A39943; I55237; A35233; S09400; B90567; S11

5269; A29439; I53466; A02852; I37247

R;D'Alessio, M.; Bernard, M.; Pretorius, P.J.; de Wet, W.; Ramirez, F.; Pretorius, P.J.

Gene 67, 105-115, 1988

A;Title: Complete nucleotide sequence of the region encompassing the first twenty-five e.

A;Reference number: I60114; MUID:88329734; PMID:2843432

A;Accession: I60114

A;Status: translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-369, 'L', 371-589 <DAL>

A;Cross-references: UNIPROT:P02452; UNIPROT:Q14992; UNIPROT:Q16053; UNIPROT:Q13896; UNIP.

R;Tromp, G.; Kuivaniemi, H.; Stacey, A.; Shikata, H.; Baldwin, C.T.; Jaenisch, R.; Prock

Biochem. J. 253, 919-922, 1988

A;Title: Structure of a full-length cDNA clone for the prepro-alpha-1(I) chain of human

A;Reference number: S01143; MUID:89025644; PMID:3178743

A;Accession: S01143

A;Molecule type: mRNA

A;Residues: 1-472 <TRO>

A;Cross-references: UNIPARC:UPI000016A6F9; EMBL:X07884; NID:g30015; PIDN:CAA30731.1; PID

A;Note: submitted to the EMBL/GenBank/DBJ databases by Prockop, D.J., 13-JUN-1988

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ding, J.F.; Morabito, M.; Myers, J.; Williams, C.;

Nature 310, 337-340, 1984

A;Title: Human proalpha1(I) collagen gene structure reveals evolutionary conservation of

A;Reference number: A93335; MUID:84270697; PMID:6462220

A;Accession: A93335

A;Molecule type: DNA

A;Residues: 1-58, 'Q', 60-181 <CHU>

A;Cross-references: UNIPARC:UPI0000173B3C; EMBL:X00820; NID:g35657; PIDN:CAA25394.1; PID

R;Rossouw, C.M.S.; Vergeer, W.P.; du Plooy, S.J.; Bernard, M.P.; Ramirez, F.; de Wet, W.

J. Biol. Chem. 262, 15151-15157, 1987

A;Title: DNA sequences in the first intron of the human pro-alpha 1(I) collagen gene en

A;Reference number: I55254; MUID:88033098; PMID:2822714

A;Accession: I55254

A;Status: translation not shown; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-45 <ROS>

A;Cross-references: UNIPARC:UPI000016A6B2; GB:J02829; NID:g180387; PIDN:AAA51993.1; PID:

R;Bornstein, P.; McKay, J.; Morishima, J.K.; Devarayalu, S.; Gelinas, R.E.

Proc. Natl. Acad. Sci. U.S.A. 84, 869-8873, 1987

A;Title: Regulatory elements in the first intron contribute to transcriptional control o

A;Reference number: A39943; MUID:88097389; PMID:3480516

A;Accession: A39943

A;Molecule type: DNA

A;Residues: 1-34 <BOR>

A;Cross-references: UNIPARC:UPI000016A6B1; GB:J03559; NID:g180876; PIDN:AAA52052.1; PID:

R;Chu, M.L.; de Wet, W.; Bernard, M.; Ramirez, F.

J. Biol. Chem. 260, 2315-2320, 1985
A;Title: Fine structural analysis of the human pro-alpha 1 (I) collagen gene. Promoter structure and organization of the 5' and 3' ends only are shown; mutant sequence 263-Arg report
A;Reference number: 155237; MUID:85130970; PMID:285713
A;Accession: 155237
A;Status: translation not shown; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-34 <CH2>
A;Cross-references: UNIPARC:UPI000016A6B1; GB:M10627; NID:G180383; PIDN:AAAS1992.1; PID:G180383
R;Wirtz, M.K.; Keene, D.R.; Hori, H.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, J. Biol. Chem. 265, 6312-6317, 1990
A;Title: In vivo and in vitro noncovalent association of excised alpha 1(I) amino-terminal propeptide, type VII.
A;Reference number: A35233; MUID:90202908; PMID:2318855
A;Accession: A35233
A;Molecule type: protein
A;Residues: 33-52 <WIR>
A;Cross-references: UNIPARC:UPI0000173B3D
A;Note: this propeptide fragment remained non-covalently bound to a defective, uncleaved alpha 1(I) chain of type I collagen
R;Weil, D.; d'Alessio, M.; Ramirez, F.; de Wet, W.; Cole, W.G.; Chan, D.; Bateman, J.F. EMBO J. 8, 1705-1710, 1989
A;Title: A base substitution in the exon of a collagen gene causes alternative splicing of the alpha 1(I) chain of type I collagen
A;Reference number: S09400; MUID:89356643; PMID:2767050
A;Accession: S09400
A;Molecule type: mRNA
A;Residues: 156-183 <WEI>
A;Cross-references: UNIPARC:UPI0000173B3E
R;Click, E.M.; Bornstein, P. Biochemistry 9, 4699-4706, 1970
A;Title: Isolation and characterization of the cyanogen bromide peptides from the alpha 1(I) chain of type I collagen
A;Reference number: A90567; MUID:71038625; PMID:5529814
A;Contents: CNBr0-1, CNBr2, CNBr4, CNBr5
A;Accession: B90567
A;Molecule type: protein
A;Residues: 162-198, 'Z', 200-201, 'Z', 203-206, 'Z', 208-209, 'Z', 211-228, 'B', 230, 'BB', 233, 'Z', 235, 'Z', 237, 'Z', 240-241, 'Z', 243-244, 'Z', 246-247, 'Z', 249-250, 'Z', 252-253, 'Z', 255-256, 'Z', 258-259, 'Z', 261-262, 'Z', 264-265, 'Z', 267-268, 'Z', 270-271, 'Z', 273-274, 'Z', 276-277, 'Z', 279-280, 'Z', 282-283, 'Z', 285-286, 'Z', 288-289, 'Z', 291-292, 'Z', 294-295, 'Z', 297-298, 'Z', 300-301, 'Z', 303-304, 'Z', 306-307, 'Z', 309-310, 'Z', 312-313, 'Z', 315-316, 'Z', 318-319, 'Z', 321-322, 'Z', 324-325, 'Z', 327-328, 'Z', 330-331, 'Z', 333-334, 'Z', 336-337, 'Z', 339-340, 'Z', 342-343, 'Z', 345-346, 'Z', 348-349, 'Z', 351-352, 'Z', 354-355, 'Z', 357-358, 'Z', 360-361, 'Z', 363-364, 'Z', 366-367, 'Z', 369-370, 'Z', 372-373, 'Z', 375-376, 'Z', 378-379, 'Z', 381-382, 'Z', 384-385, 'Z', 387-388, 'Z', 390-391, 'Z', 393-394, 'Z', 396-397, 'Z', 399-400, 'Z', 402-403, 'Z', 405-406, 'Z', 408-409, 'Z', 411-412, 'Z', 414-415, 'Z', 417-418, 'Z', 420-421, 'Z', 423-424, 'Z', 426-427, 'Z', 429-430, 'Z', 432-433, 'Z', 435-436, 'Z', 438-439, 'Z', 441-442, 'Z', 444-445, 'Z', 447-448, 'Z', 450-451, 'Z', 453-454, 'Z', 456-457, 'Z', 459-460, 'Z', 462-463, 'Z', 465-466, 'Z', 468-469, 'Z', 471-472, 'Z', 474-475, 'Z', 477-478, 'Z', 480-481, 'Z', 483-484, 'Z', 486-487, 'Z', 489-490, 'Z', 492-493, 'Z', 495-496, 'Z', 498-499, 'Z', 501-502, 'Z', 504-505, 'Z', 507-508, 'Z', 510-511, 'Z', 513-514, 'Z', 516-517, 'Z', 519-520, 'Z', 522-523, 'Z', 524-525, 'Z', 527-528, 'Z', 530-531, 'Z', 533-534, 'Z', 536-537, 'Z', 539-540, 'Z', 542-543, 'Z', 545-546, 'Z', 548-549, 'Z', 551-552, 'Z', 554-555, 'Z', 557-558, 'Z', 560-561, 'Z', 563-564, 'Z', 566-567, 'Z', 569-570, 'Z', 572-573, 'Z', 575-576, 'Z', 578-579, 'Z', 581-582, 'Z', 584-585, 'Z', 587-588, 'Z', 590-591, 'Z', 593-594, 'Z', 596-597, 'Z', 599-600, 'Z', 602-603, 'Z', 605-606, 'Z', 608-609, 'Z', 611-612, 'Z', 614-615, 'Z', 617-618, 'Z', 620-621, 'Z', 623-624, 'Z', 626-627, 'Z', 629-630, 'Z', 632-633, 'Z', 635-636, 'Z', 638-639, 'Z', 641-642, 'Z', 644-645, 'Z', 647-648, 'Z', 650-651, 'Z', 653-654, 'Z', 656-657, 'Z', 659-660, 'Z', 662-663, 'Z', 665-666, 'Z', 668-669, 'Z', 671-672, 'Z', 674-675, 'Z', 677-678, 'Z', 680-681, 'Z', 683-684, 'Z', 686-687, 'Z', 689-690, 'Z', 692-693, 'Z', 695-696, 'Z', 698-699, 'Z', 701-702, 'Z', 704-705, 'Z', 707-708, 'Z', 710-711, 'Z', 713-714, 'Z', 716-717, 'Z', 719-720, 'Z', 722-723, 'Z', 725-726, 'Z', 728-729, 'Z', 731-732, 'Z', 734-735, 'Z', 737-738, 'Z', 740-741, 'Z', 743-744, 'Z', 746-747, 'Z', 749-750, 'Z', 752-753, 'Z', 755-756, 'Z', 758-759, 'Z', 761-762, 'Z', 764-765, 'Z', 767-768, 'Z', 770-771, 'Z', 773-774, 'Z', 776-777, 'Z', 779-780, 'Z', 782-783, 'Z', 785-786, 'Z', 788-789, 'Z', 791-792, 'Z', 794-795, 'Z', 797-798, 'Z', 800-801, 'Z', 803-804, 'Z', 806-807, 'Z', 809-810, 'Z', 812-813, 'Z', 815-816, 'Z', 818-819, 'Z', 821-822, 'Z', 824-825, 'Z', 827-828, 'Z', 830-831, 'Z', 833-834, 'Z', 836-837, 'Z', 839-840, 'Z', 842-843, 'Z', 845-846, 'Z', 848-849, 'Z', 851-852, 'Z', 854-855, 'Z', 857-858, 'Z', 860-861, 'Z', 863-864, 'Z', 866-867, 'Z', 869-870, 'Z', 872-873, 'Z', 875-876, 'Z', 878-879, 'Z', 881-882, 'Z', 884-885, 'Z', 887-888, 'Z', 890-891, 'Z', 893-894, 'Z', 896-897, 'Z', 899-900, 'Z', 902-903, 'Z', 905-906, 'Z', 908-909, 'Z', 911-912, 'Z', 914-915, 'Z', 917-918, 'Z', 920-921, 'Z', 923-924, 'Z', 926-927, 'Z', 929-930, 'Z', 932-933, 'Z', 935-936, 'Z', 938-939, 'Z', 941-942, 'Z', 944-945, 'Z', 947-948, 'Z', 950-951, 'Z', 953-954, 'Z', 956-957, 'Z', 959-960, 'Z', 962-963, 'Z', 965-966, 'Z', 968-969, 'Z', 971-972, 'Z', 974-975, 'Z', 977-978, 'Z', 980-981, 'Z', 983-984, 'Z', 986-987, 'Z', 989-990, 'Z', 992-993, 'Z', 995-996, 'Z', 998-999, 'Z', 1000-1001, 'Z', 1002-1003, 'Z', 1005-1006, 'Z', 1008-1009, 'Z', 1011-1012, 'Z', 1014-1015, 'Z', 1017-1018, 'Z', 1020-1021, 'Z', 1023-1024, 'Z', 1026-1027, 'Z', 1029-1030, 'Z', 1032-1033, 'Z', 1035-1036, 'Z', 1038-1039, 'Z', 1041-1042, 'Z', 1044-1045, 'Z', 1047-1048, 'Z', 1050-1051, 'Z', 1053-1054, 'Z', 1056-1057, 'Z', 1059-1060, 'Z', 1062-1063, 'Z', 1065-1066, 'Z', 1068-1069, 'Z', 1071-1072, 'Z', 1074-1075, 'Z', 1077-1078, 'Z', 1080-1081, 'Z', 1083-1084, 'Z', 1086-1087, 'Z', 1089-1090, 'Z', 1092-1093, 'Z', 1095-1096, 'Z', 1098-1099, 'Z', 1101-1102, 'Z', 1104-1105, 'Z', 1107-1108, 'Z', 1110-1111, 'Z', 1113-1114, 'Z', 1116-1117, 'Z', 1119-1120, 'Z', 1122-1123, 'Z', 1125-1126, 'Z', 1128-1129, 'Z', 1131-1132, 'Z', 1134-1135, 'Z', 1137-1138, 'Z', 1140-1141, 'Z', 1143-1144, 'Z', 1146-1147, 'Z', 1149-1150, 'Z', 1152-1153, 'Z', 1155-1156, 'Z', 1158-1159, 'Z', 1161-1162, 'Z', 1164-1165, 'Z', 1167-1168, 'Z', 1170-1171, 'Z', 1173-1174, 'Z', 1176-1177, 'Z', 1179-1180, 'Z', 1182-1183, 'Z', 1185-1186, 'Z', 1188-1189, 'Z', 1191-1192, 'Z', 1194-1195, 'Z', 1197-1198, 'Z', 1200-1201, 'Z', 1203-1204, 'Z', 1206-1207, 'Z', 1209-1210, 'Z', 1212-1213, 'Z', 1215-1216, 'Z', 1218-1219, 'Z', 1221-1222, 'Z', 1224-1225, 'Z', 1227-1228, 'Z', 1230-1231, 'Z', 1233-1234, 'Z', 1236-1237, 'Z', 1239-1240, 'Z', 1242-1243, 'Z', 1245-1246, 'Z', 1248-1249, 'Z', 1251-1252, 'Z', 1254-1255, 'Z', 1257-1258, 'Z', 1260-1261, 'Z', 1263-1264, 'Z', 1266-1267, 'Z', 1269-1270, 'Z', 1272-1273, 'Z', 1275-1276, 'Z', 1278-1279, 'Z', 1281-1282, 'Z', 1284-1285, 'Z', 1287-1288, 'Z', 1290-1291, 'Z', 1293-1294, 'Z', 1296-1297, 'Z', 1299-1300, 'Z', 1302-1303, 'Z', 1305-1306, 'Z', 1308-1309, 'Z', 1311-1312, 'Z', 1314-1315, 'Z', 1317-1318, 'Z', 1320-1321, 'Z', 1323-1324, 'Z', 1326-1327, 'Z', 1329-1330, 'Z', 1332-1333, 'Z', 1335-1336, 'Z', 1338-1339, 'Z', 1341-1342, 'Z', 1344-1345, 'Z', 1347-1348, 'Z', 1350-1351, 'Z', 1353-1354, 'Z', 1356-1357, 'Z', 1359-1360, 'Z', 1362-1363, 'Z', 1365-1366, 'Z', 1368-1369, 'Z', 1371-1372, 'Z', 1374-1375, 'Z', 1377-1378, 'Z', 1380-1381, 'Z', 1383-1384, 'Z', 1386-1387, 'Z', 1389-1390, 'Z', 1392-1393, 'Z', 1395-1396, 'Z', 1398-1399, 'Z', 1401-1402, 'Z', 1404-1405, 'Z', 1407-1408, 'Z', 1410-1411, 'Z', 1413-1414, 'Z', 1416-1417, 'Z', 1419-1420, 'Z', 1422-1423, 'Z', 1425-1426, 'Z', 1428-1429, 'Z', 1431-1432, 'Z', 1434-1435, 'Z', 1437-1438, 'Z', 1440-1441, 'Z', 1443-1444, 'Z', 1446-1447, 'Z', 1449-1450, 'Z', 1452-1453, 'Z', 1455-1456, 'Z', 1458-1459, 'Z', 1461-1462, 'Z', 1464-1465, 'Z', 1467-1468, 'Z', 1470-1471, 'Z', 1473-1474, 'Z', 1476-1477, 'Z', 1479-1480, 'Z', 1482-1483, 'Z', 1485-1486, 'Z', 1488-1489, 'Z', 1491-1492, 'Z', 1494-1495, 'Z', 1497-1498, 'Z', 1500-1501, 'Z', 1503-1504, 'Z', 1506-1507, 'Z', 1509-1510, 'Z', 1512-1513, 'Z', 1515-1516, 'Z', 1518-1519, 'Z', 1521-1522, 'Z', 1524-1525, 'Z', 1527-1528, 'Z', 1530-1531, 'Z', 1533-1534, 'Z', 1536-1537, 'Z', 1539-1540, 'Z', 1542-1543, 'Z', 1545-1546, 'Z', 1548-1549, 'Z', 1551-1552, 'Z', 1554-1555, 'Z', 1557-1558, 'Z', 1560-1561, 'Z', 1563-1564, 'Z', 1566-1567, 'Z', 1569-1570, 'Z', 1572-1573, 'Z', 1575-1576, 'Z', 1578-1579, 'Z', 1581-1582, 'Z', 1584-1585, 'Z', 1587-1588, 'Z', 1590-1591, 'Z', 1593-1594, 'Z', 1596-1597, 'Z', 1599-1600, 'Z', 1602-1603, 'Z', 1605-1606, 'Z', 1608-1609, 'Z', 1611-1612, 'Z', 1614-1615, 'Z', 1617-1618, 'Z', 1620-1621, 'Z', 1623-1624, 'Z', 1626-1627, 'Z', 1629-1630, 'Z', 1632-1633, 'Z', 1635-1636, 'Z', 1638-1639, 'Z', 1641-1642, 'Z', 1644-1645, 'Z', 1647-1648, 'Z', 1650-1651, 'Z', 1653-1654, 'Z', 1656-1657, 'Z', 1659-1660, 'Z', 1662-1663, 'Z', 1665-1666, 'Z', 1668-1669, 'Z', 1671-1672, 'Z', 1674-1675, 'Z', 1677-1678, 'Z', 1680-1681, 'Z', 1683-1684, 'Z', 1686-1687, 'Z', 1689-1690, 'Z', 1692-1693, 'Z', 1695-1696, 'Z', 1698-1699, 'Z', 1701-1702, 'Z', 1704-1705, 'Z', 1707-1708, 'Z', 1710-1711, 'Z', 1713-1714, 'Z', 1716-1717, 'Z', 1719-1720, 'Z', 1722-1723, 'Z', 1725-1726, 'Z', 1728-1729, 'Z', 1731-1732, 'Z', 1734-1735, 'Z', 1737-1738, 'Z', 1740-1741, 'Z', 1743-1744, 'Z', 1746-1747, 'Z', 1749-1750, 'Z', 1752-1753, 'Z', 1755-1756, 'Z', 1758-1759, 'Z', 1761-1762, 'Z', 1764-1765, 'Z', 1767-1768, 'Z', 1770-1771, 'Z', 1773-1774, 'Z', 1776-1777, 'Z', 1779-1780, 'Z', 1782-1783, 'Z', 1785-1786, 'Z', 1788-1789, 'Z', 1791-1792, 'Z', 1794-1795, 'Z', 1797-1798, 'Z', 1800-1801, 'Z', 1803-1804, 'Z', 1806-1807, 'Z', 1809-1810, 'Z', 1812-1813, 'Z', 1815-1816, 'Z', 1818-1819, 'Z', 1821-1822, 'Z', 1824-1825, 'Z', 1827-1828, 'Z', 1830-1831, 'Z', 1833-1834, 'Z', 1836-1837, 'Z', 1839-1840, 'Z', 1842-1843, 'Z', 1845-1846, 'Z', 1848-1849, 'Z', 1851-1852, 'Z', 1854-1855, 'Z', 1857-1858, 'Z', 1860-1861, 'Z', 1863-1864, 'Z', 1866-1867, 'Z', 1869-1870, 'Z', 1872-1873, 'Z', 1875-1876, 'Z', 1878-1879, 'Z', 1881-1882, 'Z', 1884-1885, 'Z', 1887-1888, 'Z', 1890-1891, 'Z', 1893-1894, 'Z', 1896-1897, 'Z', 1899-1900, 'Z', 1902-1903, 'Z', 1905-1906, 'Z', 1908-1909, 'Z', 1911-1912, 'Z', 1914-1915, 'Z', 1917-1918, 'Z', 1920-1921, 'Z', 1923-1924, 'Z', 1926-1927, 'Z', 1929-1930, 'Z', 1932-1933, 'Z', 1935-1936, 'Z', 1938-1939, 'Z', 1941-1942, 'Z', 1944-1945, 'Z', 1947-1948, 'Z', 1950-1951, 'Z', 1953-1954, 'Z', 1956-1957, 'Z', 1959-1960, 'Z', 1962-1963, 'Z', 1965-1966, 'Z', 1968-1969, 'Z', 1971-1972, 'Z', 1974-1975, 'Z', 1977-1978, 'Z', 1980-1981, 'Z', 1983-1984, 'Z', 1986-1987, 'Z', 1989-1990, 'Z', 1992-1993, 'Z', 1995-1996, 'Z', 1998-1999, 'Z', 2000-2001, 'Z', 2002-2003, 'Z', 2004-2005, 'Z', 2006-2007, 'Z', 2008-2009, 'Z', 2010-2011, 'Z', 2012-2013, 'Z', 2014-2015, 'Z', 2016-2017, 'Z', 2018-2019, 'Z', 2020-2021, 'Z', 2022-2023, 'Z', 2024-2025, 'Z', 2026-2027, 'Z', 2028-2029, 'Z', 2030-2031, 'Z', 2032-2033, 'Z', 2034-2035, 'Z', 2036-2037, 'Z', 2038-2039, 'Z', 2040-2041, 'Z', 2042-2043, 'Z', 2044-2045, 'Z', 2046-2047, 'Z', 2048-2049, 'Z', 2050-2051, 'Z', 2052-2053, 'Z', 2054-2055, 'Z', 2056-2057, 'Z', 2058-2059, 'Z', 2060-2061, 'Z', 2062-2063, 'Z', 2064-2065, 'Z', 2066-2067, 'Z', 2068-2069, 'Z', 2070-2071, 'Z', 2072-2073, 'Z', 2074-2075, 'Z', 2076-2077, 'Z', 2078-2079, 'Z', 2080-2081, 'Z', 2082-2083, 'Z', 2084-2085, 'Z', 2086-2087, 'Z', 2088-2089, 'Z', 2090-2091, 'Z', 2092-2093, 'Z', 2094-2095, 'Z', 2096-2097, 'Z', 2098-2099, 'Z', 2100-2101, 'Z', 2102-2103, 'Z', 2104-2105, 'Z', 2106-2107, 'Z', 2108-2109, 'Z', 2110-2111, 'Z', 2112-2113, 'Z', 2114-2115, 'Z', 2116-2117, 'Z', 2118-2119, 'Z', 2120-2121, 'Z', 2122-2123, 'Z', 2124-2125, 'Z', 2126-2127, 'Z', 2128-2129, 'Z', 2130-2131, 'Z', 2132-2133, 'Z', 2134-2135, 'Z', 2136-2137, 'Z', 2138-2139, 'Z', 2140-2141, 'Z', 2142-2143, 'Z', 2144-2145, 'Z', 2146-2147, 'Z', 2148-2149, 'Z', 2150-2151, 'Z', 2152-2153, 'Z', 2154-2155, 'Z', 2156-2157, 'Z', 2158-2159, 'Z', 2160-2161, 'Z', 2162-2163, 'Z', 2164-2165, 'Z', 2166-2167, 'Z', 2168-2169, 'Z', 2170-2171, 'Z', 2172-2173, 'Z', 2174-2175, 'Z', 2176-2177, 'Z', 2178-2179, 'Z', 2180-2181, 'Z', 2182-2183, 'Z', 2184-2185, 'Z', 2186-2187, 'Z', 2188-2189, 'Z', 2190-2191, 'Z', 2192-2193, 'Z', 2194-2195, 'Z', 2196-2197, 'Z', 2198-2199, 'Z', 2200-2201, 'Z', 2202-2203, 'Z', 2204-2205, 'Z', 2206-2207, 'Z', 2208-2209, 'Z', 2210-2211, 'Z', 2212-2213, 'Z', 2214-2215, 'Z', 2216-2217, 'Z', 2218-2219, 'Z', 2220-2221, 'Z', 2222-2223, 'Z', 2224-2225, 'Z', 2226-2227, 'Z', 2228-2229, 'Z', 2230-2231, 'Z', 2232-2233, 'Z', 2234-2235, 'Z', 2236-2237, 'Z', 2238-2239, 'Z', 2240-2241, 'Z', 2242-2243, 'Z', 2244-2245, 'Z', 2246-2247, 'Z', 2248-2249, 'Z', 2250-2251, 'Z', 2252-2253, 'Z', 2254-2255, 'Z', 2256-2257, 'Z', 2258-2259, 'Z', 2260-2261, 'Z', 2262-2263, 'Z', 2264-2265, 'Z', 2266-2267, 'Z', 2268-2269, 'Z', 2270-2271, 'Z', 2272-2273, 'Z', 2274-2275, 'Z', 2276-2277, 'Z', 2278-2279, 'Z', 2280-2281, 'Z', 2282-2283, 'Z', 2284-2285, 'Z', 2286-2287, 'Z', 2288-2289, 'Z', 2290-2291, 'Z', 2292-2293, 'Z', 2294-2295, 'Z', 2296-2297, 'Z', 2298-2299, 'Z', 2300-2301, 'Z', 2302-2303, 'Z', 2304-2305, 'Z', 2306-2307, 'Z', 2308-2309, 'Z', 2310-2311, 'Z', 2312-2313, 'Z', 2314-2315, 'Z', 2316-2317, 'Z', 2318-2319, 'Z', 2320-2321, 'Z', 2322-2323, 'Z', 2324-2325, 'Z', 2326-2327, 'Z', 2328-2329, 'Z', 2330-2331, 'Z', 2332-2333, 'Z', 2334-2335, 'Z', 2336-2337, 'Z', 2338-2339, 'Z', 2340-2341, 'Z', 2342-2343, 'Z', 2344-2345, 'Z', 2346-2347, 'Z', 2348-2349, 'Z', 2350-2351, 'Z', 2352-2353, 'Z', 2354-2355, 'Z', 2356-2357, 'Z', 2358-2359, 'Z', 2360-2361, 'Z', 2362-2363, 'Z', 2364-2365, 'Z', 2366-2367, 'Z', 2368-2369, 'Z', 2370-2371, 'Z', 2372-2373, 'Z', 2374-2375, 'Z', 2376-2377, 'Z', 2378-2379, 'Z', 2380-2381, 'Z', 2382-2383, 'Z', 2384-2385, 'Z', 2386-2387, 'Z', 2388-2389, 'Z', 2390-2391, 'Z', 2392-2393, 'Z', 2394-2395, 'Z', 2396-2397, 'Z', 2398-2399, 'Z', 2400-2401, 'Z', 2402-2403, 'Z', 2404-2405, 'Z', 2406-2407, 'Z', 2408-2409, 'Z', 2410-2411, 'Z', 2412-2413, 'Z', 2414-2415, 'Z', 2416-2417, 'Z', 2418-2419, 'Z', 2420-2421, 'Z', 2422-2423, 'Z', 2424-2425, 'Z', 2426-2427, 'Z', 2428-2429, 'Z', 2430-2431, 'Z', 2432-2433, 'Z', 2434-2435, 'Z', 2436-2437, 'Z', 2438-2439, 'Z', 2440-2441, 'Z', 2442-2443, 'Z', 2444-2445, 'Z', 2446-2447, 'Z', 2448-2449, 'Z', 2450-2451, 'Z', 2452-2453, 'Z', 2454-2455, 'Z', 2456-2457, 'Z', 2458-2459, 'Z', 2460-2461, 'Z', 2462-2463, 'Z', 2464-2465, 'Z', 2466-2467, 'Z', 2468-2469, 'Z', 2470-2471, 'Z', 2472-2473, 'Z', 2474-2475, 'Z', 2476-2477, 'Z', 2478-2479, 'Z', 2480-2481, 'Z', 2482-2483, 'Z', 2484-2485, 'Z', 2486-2487, 'Z', 2488-2489, 'Z', 2490-2491, 'Z', 2492-2493, 'Z', 2494-2495, 'Z', 2496-2497, 'Z', 2498-2499, 'Z', 2500-2501, 'Z', 2502-2503, 'Z', 2504-2505, 'Z', 2506-2507, 'Z', 2508-2509, 'Z', 2510-2511, 'Z', 2512-2513,

A;Experimental source: fetal cell 86-146
A;Accession: E47426
A;Molecule type: mRNA
A;Residues: 1179-1387, 'R', 1389-1464 <CH7>
A;Cross-references: UNIPARC:UPI0000173B4F
A;Experimental source: fetal cell 88-251
R;Cohn, D.H.; Apone, S.; Eyre, D.R.; Starman, B.J.; Andreassen, P.; Charbonneau, H.; Nic
J. Biol. Chem. 263, 14605-14607, 1988

Query Match 98.6%; Score 1310; DB 1; Length 1464;
Best Local Similarity 98.8%; Pred. No. 1.7e-105;
Matches 242; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 3 DANVVRDLVDVDTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWIIDPN 62
DB 1220 DANVVRDLVDVDTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWIIDPN 1279
QY 63 QGCNLDIAKVFCCNMETGETCVYPTQPSVAQKNWYISKNPKDKRHVWFGEISMTDGFQPEYG 122
DB 1280 QGCNLDIAKVFCCNMETGETCVYPTQPSVAQKNWYISKNPKDKRHVWFGEISMTDGFQPEYG 1339
QY 123 QGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
DB 1340 QGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1399
QY 183 AEGNSRTYTVVDCGCTSHTGAMGKTVIEYTKKSRPLIIDVAPLDVGAPDQEFQPDVG 242
DB 1400 AEGNSRTYTVVDCGCTSHTGAMGKTVIEYTKKSRPLIIDVAPLDVGAPDQEFQPDVG 1459
QY 243 PVCFL 247
DB 1460 PVCFL 1464

RESULT 2
150629
collagen - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
A;Reference number: 150629
A;Accession: 150629
A;Molecule type: mRNA
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Residues: 1-473 <FUL>
A;Cross-references: UNIPROT:P02457; UNIPARC:UPI00001712B8; EMBL:V00401; NID:963307; PIDN:
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
F;244-473/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 92.2%; Score 1225; DB 2; Length 473;
Best Local Similarity 89.0%; Pred. No. 1e-98;
Matches 218; Conservative 20; Mismatches 7; Indels 0; Gaps 0;
QY 3 DANVVRDLVDVDTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWIIDPN 62
DB 229 DANVVRDLVDVDTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWIIDPN 288
QY 63 QGCNLDIAKVFCCNMETGETCVYPTQPSVAQKNWYISKNPKDKRHVWFGEISMTDGFQPEYG 122
DB 289 QGCNLDIAKVFCCNMETGETCVYPTQPSVAQKNWYISKNPKDKRHVWFGEISMTDGFQPEYG 348
QY 123 QGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
DB 349 QGSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 408
QY 183 AEGNSRTYTVVDCGCTSHTGAMGKTVIEYTKKSRPLIIDVAPLDVGAPDQEFQPDVG 242
DB 409 AEGNSRTYTVVDCGCTSHTGAMGKTVIEYTKKSRPLIIDVAPLDVGAPDQEFQPDVG 468
QY 243 PVCFL 247

Db 469 PVCFL 473

RESULT 3

S21626

Collagen alpha 1(I) chain, precursor - mouse
C;Species: Mus musculus (house mouse)

C;Date: 13-Jan-1995 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
A;Accession: S57243; S16374; A23982; I49557; S39789; I48300; S21626
R;Li, S.W.; Khillan, J.; Prockop, D.J.

Matrix Biol. 14, 593-595, 1994

A;Title: The complete cDNA coding sequence for the mouse pro-alpha-1(I) chain of type I

A;Reference number: S57243

A;Accession: S57243

A;Molecule type: mRNA

A;Residues: 1-1453 <LIS>

A;Cross-references: UNIPROT:P11087; UNIPARC:UPI0000027558; EMBL:U08020; NID:9470673; PIDN:

R;Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.

Biochim. Biophys. Acta 1089, 241-243, 1991

A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.

A;Reference number: S16176; MUID:91274355; PMID:2054384

A;Accession: S16374

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1442-1453 <NET>

A;Cross-references: UNIPARC:UPI0000000691; EMBL:X57981; NID:g50484; PIDN:CAA41046.1; PID

R;French, B.T.; Lee, W.H.; Maul, G.G.

Gene 39, 311-312, 1985

A;Title: Nucleotide sequence of a cDNA clone for mouse proalpha1(I) collagen protein.

A;Reference number: A23982; MUID:86137403; PMID:3841523

A;Accession: A23982

A;Molecule type: mRNA

A;Residues: 518-1128 <PRE>

A;Cross-references: UNIPARC:UPI000016CC46; GB:M14423; NID:g192261; PIDN:AAA37333.1; PIDN:

R;Monson, J.M.; Friedman, J.; McCarthy, B.J.

Mol. Cell. Biol. 2, 1362-1371, 1982

A;Title: DNA sequence analysis of a mouse pro-alpha-1(I) procollagen gene: Evidence for

A;Reference number: I49559; MUID:83141374; PMID:6298597

A;Accession: I49559

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 735-1130 <RES>

A;Cross-references: UNIPARC:UPI000016CC47; GB:M17491; NID:g192263; PIDN:AAA37334.1; PIDN:

R;Harbers, K.; Kuehn, M.; Delius, H.; Jaenisch, R.

Proc. Natl. Acad. Sci. U.S.A. 81, 1504-1508, 1984

A;Title: Insertion of retrovirus into the first intron of alpha1(I) collagen gene leads

A;Reference number: I49557; MUID:84170331; PMID:6324198

A;Accession: I49557

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-25 <RES2>

A;Cross-references: UNIPARC:UPI000005B79; GB:K01688; NID:g192246; PIDN:AAA37330.1; PIDN:

R;Fenton, S.P.; Laman, S.R.; Hannagan, M.; Stacey, A.; Jaenisch, R.; Bateman, J.F.

Biochim. Biophys. Acta 1216, 469-474, 1993

A;Title: Genomic sequence of mouse COL1A1 encoding the collagen propeptides.

A;Reference number: S39789; MUID:94092741; PMID:8368229

A;Accession: S39789

A;Molecule type: DNA

A;Residues: 1-80, 'E', 82-105, 'D', 107-185, 1031-1201, 'G', 1203-1218, 'E', 1220-1221, 'T', 1223-1

A;Cross-references: UNIPARC:UPI000005B655; UNIPARC:UPI000017738A

R;Rhodes, K.; Rippe, R.A.; Umegawa, A.; Nehls, M.; Brenner, D.A.; Breindl, M.

Mol. Cell. Biol. 14, 5950-5960, 1994

A;Title: DNA methylation represses the murine alpha 1(I) collagen promoter by an indirect

A;Reference number: I48300; MUID:94344105; PMID:8065328

A;Accession: I48300

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-80, 'E', 82-105, 'D', 107-147 <REF>

A;Cross-references: UNIPARC:UPI000017738B; EMBL:X54876; NID:g50486; PIDN:CAA38657.1; PID

C;Genetics: COL1A1

A;Introns: 770/3; 788/3; 806/3; 842/3; 860/3; 878/3; 932/3; 968/3; 1004/3; 1022/3; 1058/3;

A;Residues: 1-8,'T','10-28 <NUN>
A;Cross-references: UNIPARC:UPI000016A71A; GB:M25698; NID:g180872; PIDN:AAAS2051.1; PID:
R;Baldwin, C.T.; Reginato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.
Biochem. J. 262, 521-528, 1989
A;Title: Structure of cDNA clones coding for human type II procollagen. The alpha-1(II)
A;Reference number: S06496; MUID:90026318; PMID:2803268
A;Accession: S06496
A;Molecule type: mRNA
A;Residues: 7-28,'R','99-157','P','159-440','G','442-456','E','458-640','A','642-831','PA','834','F'
A;Cross-references: UNIPARC:UPI0000173B65; EMBL:X16711; NID:g30040; PIDN:CAA34683.1; PID:
R;Ryan, M.C.; Sandelli, L.J.
J. Biol. Chem. 265, 10334-10339, 1990
A;Title: Differential expression of a cysteine-rich domain in the amino-terminal propept
A;Reference number: A35428; MUID:90285153; PMID:2335003
A;Accession: A35428
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 27-81,'L','83-103 <RYA2>
A;Cross-references: UNIPARC:UPI0000173B66
A;Note: alternative splice form 2; splicing appears to be under developmental regulation
R;Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.
Genomics 4, 438-441, 1989
A;Title: Organization of the exons coding for Pro alpha-1(II) collagen N-propeptide conf
A;Reference number: A30147; MUID:89233138; PMID:2714801
A;Accession: A30147
A;Molecule type: DNA
A;Residues: 104-157,'P','159-236 <SUM>
A;Cross-references: UNIPARC:UPI0000173B67; GB:J03065; GB:M23660; GB:M25655; GB:M25656; G
R;Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 6565-6568, 1990
A;Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri
A;Reference number: A94227; MUID:90370826; PMID:1975593
A;Accession: A33116
A;Molecule type: DNA
A;Residues: 171-172,'C','174-175 <ALA>
A;Cross-references: UNIPARC:UPI0000173B68
A;Note: mutant sequence from a family with family with primary generalized osteoarthritis
R;Diab, M.; Wu, J.J.; Eyre, D.R.
Biochem. J. 314, 327-332, 1996
A;Title: Collagen type IX from human cartilage: a structural profile of intermolecular c
A;Reference number: S64673; MUID:96195147; PMID:8660302
A;Accession: S64674
A;Molecule type: protein
A;Residues: 188-189,'X','191-195;1224-1230,'X','1232-1236 <DIA>
A;Cross-references: UNIPARC:UPI0000173B69; UNIPARC:UPI0000173B6A
R;Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,
Eur. J. Biochem. 234, 125-131, 1995
A;Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil can
A;Reference number: S63514; MUID:96096730; PMID:8529631
A;Accession: S63514
A;Molecule type: protein
A;Residues: 243-261;575-590;756-763,'X','765-779 <PRA>
A;Cross-references: UNIPARC:UPI0000173B6B; UNIPARC:UPI0000173B6C; UNIPARC:UPI0000173B6D
R;Tiller, G.E.; Weis, M.A.; Polunbo, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre,
Am. J. Hum. Genet. 56, 388-395, 1995
A;Title: An RNA-splicing mutation (G>SIVS20) in the type II collagen gene (COL2A1) in a
A;Reference number: I38867; MUID:95150028; PMID:7847372
A;Accession: I38867
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 440,'G','442-456','E','458-480','P','482-509 <TIL1>
A;Cross-references: UNIPARC:UPI000006F3AF; EMBL:U15195; NID:g557053; PIDN:AAB60370.1; PT
R;Ramirez, F.
submitted to the EMBL Data Library, December 1988
A;Reference number: S04892
A;Accession: S04892
A;Molecule type: mRNA
A;Residues: 501-676,'A','678-783,'A','785-831','PA','834','F','836-1214 <RAM>
R;Vikkula, M.; Peltonen, L.
FEBS Lett. 250, 171-174, 1989
A;Title: Structural analyses of the polymorphic area in type II collagen gene.

A;Reference number: S05000; MUID:89325561; PMID:2753125
A;Accession: S05000
A;Molecule type: DNA
A;Residues: 630-640,'A','642-785 <VIK2>
A;Cross-references: UNIPARC:UPI0000173B6E; EMBL:X16158; NID:g29951; PIDN:CAA34278.1; PID:
A;282.1; PID:g1335022; PIDN:CAA34283.1; PID:g1335023; PIDN:CAA34284.1; PID:g1335024
R;Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyre, D.
J. Biol. Chem. 267, 22522-22526, 1992
A;Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro
A;Reference number: A44309; MUID:93054548; PMID:1429602
A;Accession: A44309
A;Status: nucleic acid sequence not shown; not compared with conceptual translation
A;Molecule type: DNA; mRNA
A;Residues: 752-831,'PA','834','F','836-1005','K','1007-1036','Q','1038-1052','E','1054-1068','T'
A;Cross-references: UNIPARC:UPI0000173B6F; GB:I00977; NID:g180812; PIDN:AAB23914.1; PID:
A;Note: sequence extracted from NCBI backbone (NCBI:P:117273); parts of this sequence wer
A;Note: this translation is not annotated and this publication is not cited in GenBank e
A;Note: mutant sequence associated with perinatal lethal hypochondrogenesis
R;Tiller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.
Proc. Natl. Acad. Sci. U.S.A. 87, 3889-3893, 1990
A;Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an individua
A;Reference number: S16502; MUID:90251662; PMID:2339128
A;Accession: S16502
A;Molecule type: DNA
A;Residues: 1164-1184,'GPSKGDGANGIPGP','1185-1199 <TIL2>
A;Cross-references: UNIPARC:UPI000011F72; EMBL:M37125; NID:g180808; PIDN:AAAS2037.1; PID:
A;Note: mutant sequence from a patient with spondyloepiphyseal dysplasia
R;Cheah, K.S.E.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.
Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985
A;Title: Identification and characterization of the human type II collagen gene (COL2A1)
A;Reference number: A02858; MUID:85190534; PMID:3857598
A;Accession: A02858
A;Molecule type: DNA
A;Residues: 1032-1056,'N','1058-1068','T','1070-1487 <CHE>
A;Cross-references: UNIPARC:UPI000016A6BB; GB:J00116; NID:g180395; PIDN:AAA51997.1; PID:
R;Elima, K.; Vuorio, T.; Vuorio, E.
Nucleic Acids Res. 15, 9499-9504, 1987
A;Title: Determination of the single polyadenylation site of the human pro-alpha-1(II) c
A;Reference number: A27280; MUID:88067771; PMID:2825137
A;Accession: A27280
A;Molecule type: DNA; mRNA
A;Residues: 1175-1487 <ELI>
A;Cross-references: UNIPARC:UPI000016A71B; EMBL:X06268; NID:g30096; PIDN:CAA29604.1; PID:
A;Experimental source: fetal epiphyseal cartilage
R;van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.
Biochem. J. 237, 923-925, 1986
A;Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.
A;Reference number: A57033; MUID:87099927; PMID:3800925
A;Accession: A57033
A;Molecule type: protein
A;Residues: 'XE',1244-1246,'N',1248,'X',1250-1265;1295-1305;1395-1408 <VAN>
A;Cross-references: UNIPARC:UPI0000173B70; UNIPARC:UPI0000173B71; UNIPARC:UPI0000173B72
A;Note: chondrocalcin identified as released collagen 1(II) chain carboxyl-terminal prop
R;Strom, C.M.; Upholt, W.B
Nucleic Acids Res. 12, 1025-1038, 1984
A;Title: Isolation and characterization of genomic clones corresponding to the human typ
A;Reference number: A21733; MUID:84118798; PMID:6320112
A;Accession: A21733
A;Molecule type: DNA
A;Residues: 1245-1295 <STRI>
A;Cross-references: UNIPARC:UPI000016A61A; EMBL:X00339; EMBL:X00298; NID:g394699; PIDN:CA
A;Accession: B21733
A;Molecule type: DNA
A;Residues: 894-909,'PE' <STR2>
A;Cross-references: UNIPARC:UPI000006EA4D; GB:K01785; NID:g30035; PIDN:CAA25082.1; PID:g1
R;Nunez, A.M.; Francomano, C.; Young, M.F.; Martin, G.R.; Yamada, Y.
Biochemistry 24, 6343-6348, 1985
A;Title: Isolation and partial characterization of genomic clones coding for a human pro-
A;Reference number: A24561; MUID:86104139; PMID:3002437
A;Accession: A24561
A;Molecule type: DNA
A;Residues: 1296-1358 <NUN2>

Best Local Similarity 64.3%; Pred. No. 2.4e-72;		Matches 161; Conservative 47; Mismatches 36; Indels 4; Gaps 3;	
QY	1	RSANVVRDLVDVTTLSLSQOIEI-IRSPGSRKNPARTCRDLKMHCHSWKSGEYWI	59
DB	1248	QASSSVLQTADVEAN--LKTANNQIESIIRSPDGTCKNPARTCRDLKLCHPEWKSGEYWI	1305
QY	60	DPNOGNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTGQFQF	119
DB	1306	DPNEGCTVDAIKVFCNMETGESCVYFSPKIPKKNWWSAKG-KEKKHIWFGETINGGQFQF	1364
QY	120	EYGGQSDPADVAIQLTFLRLMSTEAQNITYHCNSVAYMDQQTGNLKKALLKGSNEI	179
DB	1365	SYGDDSSAPNTANIQLTFLRLSTDSQNITYHCNSIAPMDRASGNLKKAVILQGSNDV	1424
QY	180	EIRAEGRSRTYSVTDGCTSHGTGAKGKTVIEYTKTSRLPIIDVAPLDVGAPDQBFQF	239
DB	1425	EIRAEGRSRTYNALEDGCKKHTGKWSKTVIEYRTQKTSRLPIVDIAPMDIGGADQBFQF	1484
QY	240	DVGPCVCF 247	
DB	1485	DIGPCVCF 1492	
RESULT 12			
A61396			
collagen alpha 1(II) chain - golden hamster (fragments)			
C/Species: Mesocricetus auratus (golden hamster)			
C/Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 09-Jul-2004			
C/Accession: A61396			
R/Cizdziel, P.E.; Hosoi, J.; Montgomery, J.C.; Wiseman, R.W.; Barrett, J.C.			
Mol. Carcinog. 4, 14-24, 1991			
A/Title: Loss of a tumor suppressor gene function is correlated with downregulation of			
A/Reference number: A61396; MUID:91182265; PMID:2009131			
A/Accession: A61396			
A/Molecule type: mRNA			
A/Status: preliminary; not compared with conceptual translation			
A/Residues: 1-323 <CIZ>			
A/Cross-references: UNIPROT:Q7M099; UNIPARC:UPI0000177383			
C/Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;			
F;120-323/Domain: fibrillar collagen carboxyl-terminal homology <FC>			
Query Match 66.4%; Score 883; DB 2; Length 323;			
Best Local Similarity 64.3%; Pred. No. 3.5e-69;			
Matches 155; Conservative 35; Mismatches 25; Indels 26; Gaps 2;			
QY	7	VRDRLEVDVTTLSLSQOIEIIRSPGSRKNPARTCRDLKMHCHSWKSGEYWI	66
DB	109	LQHQHVEVDATLSLNNQIESIRSPDGSCKNPARTCQDLKLCHPEWKSQGYWIDPNQGC	168
QY	67	LDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEGGQGS	126
DB	169	LDAMKAFCNMETGETCVYPPNPASVPKKNWWSKG-KEKKHI-----	208
QY	127	DPADVAIQLTFLRLMSTEAQNITYHCNSVAYMDQQTGNLKKALLKGSNEI	186
DB	209	-----VQMTFLRLSTDSQNITYHCNSIAYLDEAAGNLKALLIQGSNDVEMRAGN	262
QY	187	SRTYSVTDGCTSHGTGAKGKTVIEYTKTSRLPIIDVAPLDVGAPDQBFQFQF	246
DB	263	SRTYTTALKDGTGKHTGKWKTVIEYRSQTSRLPIIDIAPMDIGGPEQFQFQF	322
QY	247	L 247	
DB	323	L 323	
RESULT 13			
CGGH2S			
collagen alpha 2(I) chain precursor - chicken (fragments)			
C/Species: Gallus gallus (chicken)			
C/Date: 24-Apr-1984 #sequence_revision 15-Aug-1997 #text_change 09-Jul-2004			
C/Accession: I50206; I50207; S07354; S10848; S11146; I50628; I50170; I50625; I50			

R;Boedtker, H.; Finer, M.; Aho, S.	
Ann. N. Y. Acad. Sci. 460, 85-116, 1985	
A/Title: The structure of the chicken alpha 2 collagen gene.	
A/Reference number: I50206; MUID:86185168; PMID:3868961	
A/Accession: I50206	
A/Status: translated from GB/EMBL/DDBJ	
A/Molecule type: DNA	
A/Residues: 1-245 <BOE>	
A/Cross-references: UNIPROT:P02467; UNIPARC:UPI00001712EA; GB:M25963; NID:g211581; PIDN:	
A/Accession: I50207	
A/Status: translated from GB/EMBL/DDBJ	
A/Molecule type: DNA	
A/Residues: 246-431 <BO2>	
A/Cross-references: UNIPARC:UPI00001712AA; GB:M25965; NID:g211583; PIDN:AAA69961.1; PID:	
R;Aho, S.; Tate, V.; Boedtker, H.	
Nucleic Acids Res. 12, 6117-6125, 1984	
A/Title: Location of the 11 bp exon in the chicken pro alpha-2(I) collagen gene.	
A/Reference number: S07354; MUID:84297217; PMID:6473103	
A/Accession: S07354	
A/Molecule type: DNA	
A/Residues: 1-33 <AHO>	
A/Cross-references: UNIPARC:UPI0000173BAC; EMBL:X00760; NID:g63266; PIDN:CAA25330.1; PID:	
R;Tate, V.E.; Finer, M.H.; Boedtker, H.; Doty, P.	
Nucleic Acids Res. 11, 91-104, 1983	
A/Title: Chick pro-alpha-2 (I) collagen gene: exon location and coding potential for the	
A/Reference number: S10480; MUID:83246518; PMID:6135195	
A/Accession: S10848	
A/Molecule type: mRNA	
A/Residues: 1-89 <TAT>	
A/Cross-references: UNIPARC:UPI00001712EC; EMBL:X02657; NID:g63314; PIDN:CAA26493.1; PID:	
A/Accession: S10480	
A/Molecule type: DNA	
A/Residues: 17-73 <TAW>	
A/Cross-references: UNIPARC:UPI0000173BAD	
A/Note: the authors translated the codon CAG for residue 42 as Glu	
R;Vogeli, G.; Ohkubo, H.; Sobel, M.E.; Yamada, Y.; Pastan, I.; de Crombrughe, B.	
Proc. Natl. Acad. Sci. U.S.A. 78, 5334-5338, 1981	
A/Title: Structure of the promoter for chicken alpha-2 type I collagen gene.	
A/Reference number: S11146; MUID:82060240; PMID:6946474	
A/Accession: S11146	
A/Status: preliminary	
A/Molecule type: DNA	
A/Residues: 1-15 <VOG>	
A/Cross-references: UNIPARC:UPI0000173BAE; EMBL:J00821	
R;Yamada, Y.; Avvedimento, V.E.; Mudryj, M.; Ohkubo, H.; Vogeli, G.; Irani, M.; Pastan, J.	
Cell 22, 887-892, 1980	
A/Title: The collagen gene: evidence for its evolutionary assembly by amplification of a	
A/Reference number: I50170; MUID:8112157; PMID:7460017	
A/Accession: I50628	
A/Status: preliminary; translated from GB/EMBL/DDBJ	
A/Molecule type: DNA	
A/Residues: 126-161 <YAM>	
A/Cross-references: UNIPARC:UPI00001712A9; EMBL:V00400; NID:g63305; PID:g833611	
A/Accession: I50170	
A/Status: preliminary; translated from GB/EMBL/DDBJ	
A/Molecule type: DNA	
A/Residues: 126-161 <YA2>	
A/Cross-references: UNIPARC:UPI00001712A9; GB:J00828; NID:g211295; PIDN:AAA51612.1; PID:	
A/Accession: I50625	
A/Status: preliminary; translated from GB/EMBL/DDBJ	
A/Molecule type: DNA	
A/Residues: 530-558 <YA3>	
A/Cross-references: UNIPARC:UPI0000173BAF; EMBL:V00396; NID:g63295; PID:g833609	
A/Accession: I50626	
A/Status: preliminary; translated from GB/EMBL/DDBJ	
A/Molecule type: DNA	
A/Residues: 572-601 <YA4>	
A/Cross-references: UNIPARC:UPI0000173BB0; EMBL:V00398; NID:g63299; PID:g833610	
A/Accession: I50624	
A/Status: preliminary; translated from GB/EMBL/DDBJ	
A/Molecule type: DNA	
A/Residues: 433-483 <YA5>	
A/Cross-references: UNIPARC:UPI0000001087; EMBL:V00394; NID:g63287; PID:g833608	

A;Residues: 536-558 <NV>
A;Cross-references: UNIPARC:UPI00001712AD; GB:M10581; NID:G911323; PIDN:AAA48637.1.; PID:
R;Lehrach, H.; Fritschauf, A.M.; Hanahan, D.; Wozney, J.; Fuller, F.; Crkvenjakov, R.; Bo
Proc. Natl. Acad. Sci. U.S.A. 75, 5417-5421, 1978
A;Title: Construction and characterization of a 2.5-kilobase procollagen clone.
A;Reference number: I50171; MUID:79074829; PMID:364479
A;Accession: I50171
A;Status: preliminary; translated from GB/EMBL/DBBJ
A;Molecule type: mRNA
A;Residues: 484-505 <LEH>
A;Cross-references: UNIPARC:UPI00001712AB; GB:J00837; NID:G9530617; PIDN:AAA51614.1.; PID
C;Genetics:
A;Gene: COL1A2
A;Introns: 24/1; 27/3; 33/3; 45/3; 73/3; 92/3; 107/3; 125/3; 143/3; 161/3; 179/3; 197/3;
A;Note: the list of introns is incomplete
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
C;Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroxypy
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-77/Domain: amino-terminal propeptide #status predicted <PRO>
F;78-964/Product: collagen alpha 2(I) chain (fragments) #status predicted <MATN>
F;78-89/Region: amino-terminal nonhelical telopeptide
F;737-964/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
F;78/Modified site: blocked amino end (Gln) (in mature form) (probably pyrrolidone carb
F;83/Modified site: allysine (Lys) #status experimental
F;422,425/Modified site: 4-Hydroxyproline (Pro) #status experimental
F;866/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 66.4%; Score 883; DB 1; Length 964;
Best Local Similarity 61.4%; Pred. No. 1.4e-68;
Matches 151; Conservative 44; Mismatches 49; Indels 2; Gaps 2;

Qy 1 RSDANVVRDRDLEVDITLKSI-SQQTINIRSPESGRKNPARTCRDLKMCCHDSWKSGBYIID 60
Db :
720 RADQPSLRPKDYEDVATLKLNNQIETLLTPEGSKKNPARTCRDLRLSHPEWSGGFYIID 779

Qy 61 PNQCNLDAIKVCNMETGETCVPTQPSSVAQNKYVISKNPKDKRHVWGESSMTDGQFE 120
Db ||||| :
780 PNOCTADAIRAYCDFATGETCIHASLEDIPTKTWYVKSNPKDKKHWFGE-INGGTQFE 838

Qy 121 YGGOGSDPADVAIQLTFLRLMSTASQNTIVHKNSVAYMDQOTGNLKKALLKGNEIE 180
Db ||||| :
839 YNGEVTTKDMATQLAFMLLNPHASQNTIVHKNSAIYMDEETGNLKAIVLQGSDNVE 898

Qy 181 IRAEGRSFTYSVTVDGCTSHTGAWKTVIEYKTTKSSRLPIIDVAPLDVGAPDQDFGD 240
Db ||||| :
899 LRAESG-SRFTSVLVDGCSKNNKWGKTIIERYTNKPSRLPILDIAPLDIGGADQEFGLH 957

Qy 241 VGPVCF 246
Db : |||||
958 IGPVCF 963

RESULT 14
SI0847
collagen alpha 2(I) chain - chicken (fragment)
C;Species: Gallus gallus (chicken)
C;Date: 21-Nov-1993 #sequence_revision 26-Jul-1996 #text_change 23-May-1997
C;Accession: SI0847; S65690
F;Dickson, L.A.; Ninomiya, Y.; Bernard, M.P.; Pesciotta, D.M.; Parsons, J.; Green, G.; E
J. Biol. Chem. 256, 8407-8415, 1981
A;Title: The exon/intron structure of the 3'-region of the pro-alpha-2(I) collagen gene.
A;Reference number: SI0847; MUID:81264246; PMID:6267043
A;Accession: SI0847
A;Molecule type: DNA
A;Residues: 1-365 <DIC>
A;Cross-references: UNIPARC:UPI0000177396; EMBL:J00811
A;Note: the authors translated the codon ATA for residue 207 as Asp, AGC for residue 216
A;Accession: S65690
A;Molecule type: protein
A;Residues: 121-123, 'X', 125-134, 'X', 136-137, 'X', 139, 'XXX', 143 <DIX>
A;Cross-references: UNIPARC:UPI0000177397
C;Genetics:
A;Gene: COL1A2

J. Biol. Chem. 265, 17070-17077, 1990
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping and a severe form of Ehlers-Danlos syndrome
A;Reference number: A38303; MUID:91009133; PMID:2145268
A;Accession: A38303
A;Molecule type: mRNA
A;Residues: 861-1015 <COL>
A;Cross-references: UNIPARC:UPI00000004A1; GB:J05617; GB:M55603; GB:M59227; NID:G180878;
A;Note: a mutant sequence with 942-977 spliced out from a patient with Ehlers-Danlos syndrome
R;Mankoo, B.S.; Daigleish, R.
Nucleic Acids Res. 16, 2337, 1988
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.
A;Reference number: S02119; MUID:88189827; PMID:3357782
A;Accession: S02119
A;Status: translation not shown
A;Molecule type: mRNA
A;Residues: 950-1018, 'Y', 1020-1183, 'S', 1185-1466 <MAN>
A;Cross-references: UNIPARC:UPI0000173886; EMBL:X06700; NID:G30053; PIDN:CAA2986.1; PIDN:CAA2986.1
R;Seyer, J.M.; Kang, A.H.
Biochemistry 20, 2621-2627, 1981
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from type I collagen
A;Reference number: A90446; MUID:81208139; PMID:7016180
A;Accession: A90446
A;Molecule type: protein
A;Residues: 965-979, 'A', 981-984, 'PS', 987, 'QN', 990-1096, 'P', 1098-1152, 'AT', 1155, 'S', 1157-1158, 'S', 1159-1160, 'S', 1161-1162, 'S', 1163-1164, 'S', 1165-1166, 'S', 1167-1168, 'S', 1169-1170, 'S', 1171-1172, 'S', 1173-1174, 'S', 1175-1176, 'S', 1177-1178, 'S', 1179-1180, 'S', 1181-1182, 'S', 1183-1184, 'S', 1185-1186, 'S', 1187-1188, 'S', 1189-1190, 'S', 1191-1192, 'S', 1193-1194, 'S', 1195-1196, 'S', 1197-1198, 'S', 1199-1200, 'S', 1201-1202, 'S', 1203-1204, 'S', 1205-1206, 'S', 1207-1208, 'S', 1209-1210, 'S', 1211-1212, 'S', 1213-1214, 'S', 1215-1216, 'S', 1217-1218, 'S', 1219-1220, 'S', 1221-1222, 'S', 1223-1224, 'S', 1225-1226, 'S', 1227-1228, 'S', 1229-1230, 'S', 1231-1232, 'S', 1233-1234, 'S', 1235-1236, 'S', 1237-1238, 'S', 1239-1240, 'S', 1241-1242, 'S', 1243-1244, 'S', 1245-1246, 'S', 1247-1248, 'S', 1249-1250, 'S', 1251-1252, 'S', 1253-1254, 'S', 1255-1256, 'S', 1257-1258, 'S', 1259-1260, 'S', 1261-1262, 'S', 1263-1264, 'S', 1265-1266, 'S', 1267-1268, 'S', 1269-1270, 'S', 1271-1272, 'S', 1273-1274, 'S', 1275-1276, 'S', 1277-1278, 'S', 1279-1280, 'S', 1281-1282, 'S', 1283-1284, 'S', 1285-1286, 'S', 1287-1288, 'S', 1289-1290, 'S', 1291-1292, 'S', 1293-1294, 'S', 1295-1296, 'S', 1297-1298, 'S', 1299-1300, 'S', 1301-1302, 'S', 1303-1304, 'S', 1305-1306, 'S', 1307-1308, 'S', 1309-1310, 'S', 1311-1312, 'S', 1313-1314, 'S', 1315-1316, 'S', 1317-1318, 'S', 1319-1320, 'S', 1321-1322, 'S', 1323-1324, 'S', 1325-1326, 'S', 1327-1328, 'S', 1329-1330, 'S', 1331-1332, 'S', 1333-1334, 'S', 1335-1336, 'S', 1337-1338, 'S', 1339-1340, 'S', 1341-1342, 'S', 1343-1344, 'S', 1345-1346, 'S', 1347-1348, 'S', 1349-1350, 'S', 1351-1352, 'S', 1353-1354, 'S', 1355-1356, 'S', 1357-1358, 'S', 1359-1360, 'S', 1361-1362, 'S', 1363-1364, 'S', 1365-1366, 'S', 1367-1368, 'S', 1369-1370, 'S', 1371-1372, 'S', 1373-1374, 'S', 1375-1376, 'S', 1377-1378, 'S', 1379-1380, 'S', 1381-1382, 'S', 1383-1384, 'S', 1385-1386, 'S', 1387-1388, 'S', 1389-1390, 'S', 1391-1392, 'S', 1393-1394, 'S', 1395-1396, 'S', 1397-1398, 'S', 1399-1400, 'S', 1401-1402, 'S', 1403-1404, 'S', 1405-1406, 'S', 1407-1408, 'S', 1409-1410, 'S', 1411-1412, 'S', 1413-1414, 'S', 1415-1416, 'S', 1417-1418, 'S', 1419-1420, 'S', 1421-1422, 'S', 1423-1424, 'S', 1425-1426, 'S', 1427-1428, 'S', 1429-1430, 'S', 1431-1432, 'S', 1433-1434, 'S', 1435-1436, 'S', 1437-1438, 'S', 1439-1440, 'S', 1441-1442, 'S', 1443-1444, 'S', 1445-1446, 'S', 1447-1448, 'S', 1449-1450, 'S', 1451-1452, 'S', 1453-1454, 'S', 1455-1456, 'S', 1457-1458, 'S', 1459-1460, 'S', 1461-1462, 'S', 1463-1464, 'S', 1465-1466, 'S', 1467-1468, 'S', 1469-1470, 'S', 1471-1472, 'S', 1473-1474, 'S', 1475-1476, 'S', 1477-1478, 'S', 1479-1480, 'S', 1481-1482, 'S', 1483-1484, 'S', 1485-1486, 'S', 1487-1488, 'S', 1489-1490, 'S', 1491-1492, 'S', 1493-1494, 'S', 1495-1496, 'S', 1497-1498, 'S', 1499-1500, 'S', 1501-1502, 'S', 1503-1504, 'S', 1505-1506, 'S', 1507-1508, 'S', 1509-1510, 'S', 1511-1512, 'S', 1513-1514, 'S', 1515-1516, 'S', 1517-1518, 'S', 1519-1520, 'S', 1521-1522, 'S', 1523-1524, 'S', 1525-1526, 'S', 1527-1528, 'S', 1529-1530, 'S', 1531-1532, 'S', 1533-1534, 'S', 1535-1536, 'S', 1537-1538, 'S', 1539-1540, 'S', 1541-1542, 'S', 1543-1544, 'S', 1545-1546, 'S', 1547-1548, 'S', 1549-1550, 'S', 1551-1552, 'S', 1553-1554, 'S', 1555-1556, 'S', 1557-1558, 'S', 1559-1560, 'S', 1561-1562, 'S', 1563-1564, 'S', 1565-1566, 'S', 1567-1568, 'S', 1569-1570, 'S', 1571-1572, 'S', 1573-1574, 'S', 1575-1576, 'S', 1577-1578, 'S', 1579-1580, 'S', 1581-1582, 'S', 1583-1584, 'S', 1585-1586, 'S', 1587-1588, 'S', 1589-1590, 'S', 1591-1592, 'S', 1593-1594, 'S', 1595-1596, 'S', 1597-1598, 'S', 1599-1600, 'S', 1601-1602, 'S', 1603-1604, 'S', 1605-1606, 'S', 1607-1608, 'S', 1609-1610, 'S', 1611-1612, 'S', 1613-1614, 'S', 1615-1616, 'S', 1617-1618, 'S', 1619-1620, 'S', 1621-1622, 'S', 1623-1624, 'S', 1625-1626, 'S', 1627-1628, 'S', 1629-1630, 'S', 1631-1632, 'S', 1633-1634, 'S', 1635-1636, 'S', 1637-1638, 'S', 1639-1640, 'S', 1641-1642, 'S', 1643-1644, 'S', 1645-1646, 'S', 1647-1648, 'S', 1649-1650, 'S', 1651-1652, 'S', 1653-1654, 'S', 1655-1656, 'S', 1657-1658, 'S', 1659-1660, 'S', 1661-1662, 'S', 1663-1664, 'S', 1665-1666, 'S', 1667-1668, 'S', 1669-1670, 'S', 1671-1672, 'S', 1673-1674, 'S', 1675-1676, 'S', 1677-1678, 'S', 1679-1680, 'S', 1681-1682, 'S', 1683-1684, 'S', 1685-1686, 'S', 1687-1688, 'S', 1689-1690, 'S', 1691-1692, 'S', 1693-1694, 'S', 1695-1696, 'S', 1697-1698, 'S', 1699-1700, 'S', 1701-1702, 'S', 1703-1704, 'S', 1705-1706, 'S', 1707-1708, 'S', 1709-1710, 'S', 1711-1712, 'S', 1713-1714, 'S', 1715-1716, 'S', 1717-1718, 'S', 1719-1720, 'S', 1721-1722, 'S', 1723-1724, 'S', 1725-1726, 'S', 1727-1728, 'S', 1729-1730, 'S', 1731-1732, 'S', 1733-1734, 'S', 1735-1736, 'S', 1737-1738, 'S', 1739-1740, 'S', 1741-1742, 'S', 1743-1744, 'S', 1745-1746, 'S', 1747-1748, 'S', 1749-1750, 'S', 1751-1752, 'S', 1753-1754, 'S', 1755-1756, 'S', 1757-1758, 'S', 1759-1760, 'S', 1761-1762, 'S', 1763-1764, 'S', 1765-1766, 'S', 1767-1768, 'S', 1769-1770, 'S', 1771-1772, 'S', 1773-1774, 'S', 1775-1776, 'S', 1777-1778, 'S', 1779-1780, 'S', 1781-1782, 'S', 1783-1784, 'S', 1785-1786, 'S', 1787-1788, 'S', 1789-1790, 'S', 1791-1792, 'S', 1793-1794, 'S', 1795-1796, 'S', 1797-1798, 'S', 1799-1800, 'S', 1801-1802, 'S', 1803-1804, 'S', 1805-1806, 'S', 1807-1808, 'S', 1809-1810, 'S', 1811-1812, 'S', 1813-1814, 'S', 1815-1816, 'S', 1817-1818, 'S', 1819-1820, 'S', 1821-1822, 'S', 1823-1824, 'S', 1825-1826, 'S', 1827-1828, 'S', 1829-1830, 'S', 1831-1832, 'S', 1833-1834, 'S', 1835-1836, 'S', 1837-1838, 'S', 1839-1840, 'S', 1841-1842, 'S', 1843-1844, 'S', 1845-1846, 'S', 1847-1848, 'S', 1849-1850, 'S', 1851-1852, 'S', 1853-1854, 'S', 1855-1856, 'S', 1857-1858, 'S', 1859-1860, 'S', 1861-1862, 'S', 1863-1864, 'S', 1865-1866, 'S', 1867-1868, 'S', 1869-1870, 'S', 1871-1872, 'S', 1873-1874, 'S', 1875-1876, 'S', 1877-1878, 'S', 1879-1880, 'S', 1881-1882, 'S', 1883-1884, 'S', 1885-1886, 'S', 1887-1888, 'S', 1889-1890, 'S', 1891-1892, 'S', 1893-1894, 'S', 1895-1896, 'S', 1897-1898, 'S', 1899-1900, 'S', 1901-1902, 'S', 1903-1904, 'S', 1905-1906, 'S', 1907-1908, 'S', 1909-1910, 'S', 1911-1912, 'S', 1913-1914, 'S', 1915-1916, 'S', 1917-1918, 'S', 1919-1920, 'S', 1921-1922, 'S', 1923-1924, 'S', 1925-1926, 'S', 1927-1928, 'S', 1929-1930, 'S', 1931-1932, 'S', 1933-1934, 'S', 1935-1936, 'S', 1937-1938, 'S', 1939-1940, 'S', 1941-1942, 'S', 1943-1944, 'S', 1945-1946, 'S', 1947-1948, 'S', 1949-1950, 'S', 1951-1952, 'S', 1953-1954, 'S', 1955-1956, 'S', 1957-1958, 'S', 1959-1960, 'S', 1961-1962, 'S', 1963-1964, 'S', 1965-1966, 'S', 1967-1968, 'S', 1969-1970, 'S', 1971-1972, 'S', 1973-1974, 'S', 1975-1976, 'S', 1977-1978, 'S', 1979-1980, 'S', 1981-1982, 'S', 1983-1984, 'S', 1985-1986, 'S', 1987-1988, 'S', 1989-1990, 'S', 1991-1992, 'S', 1993-1994, 'S', 1995-1996, 'S', 1997-1998, 'S', 1999-2000, 'S', 2001-2002, 'S', 2003-2004, 'S', 2005-2006, 'S', 2007-2008, 'S', 2009-2010, 'S', 2011-2012, 'S', 2013-2014, 'S', 2015-2016, 'S', 2017-2018, 'S', 2019-2020, 'S', 2021-2022, 'S', 2023-2024, 'S', 2025-2026, 'S', 2027-2028, 'S', 2029-2030, 'S', 2031-2032, 'S', 2033-2034, 'S', 2035-2036, 'S', 2037-2038, 'S', 2039-2040, 'S', 2041-2042, 'S', 2043-2044, 'S', 2045-2046, 'S', 2047-2048, 'S', 2049-2050, 'S', 2051-2052, 'S', 2053-2054, 'S', 2055-2056, 'S', 2057-2058, 'S', 2059-2060, 'S', 2061-2062, 'S', 2063-2064, 'S', 2065-2066, 'S', 2067-2068, 'S', 2069-2070, 'S', 2071-2072, 'S', 2073-2074, 'S', 2075-2076, 'S', 2077-2078, 'S', 2079-2080, 'S', 2081-2082, 'S', 2083-2084, 'S', 2085-2086, 'S', 2087-2088, 'S', 2089-2090, 'S', 2091-2092, 'S', 2093-2094, 'S', 2095-2096, 'S', 2097-2098, 'S', 2099-2100, 'S', 2101-2102, 'S', 2103-2104, 'S', 2105-2106, 'S', 2107-2108, 'S', 2109-2110, 'S', 2111-2112, 'S', 2113-2114, 'S', 2115-2116, 'S', 2117-2118, 'S', 2119-2120, 'S', 2121-2122, 'S', 2123-2124, 'S', 2125-2126, 'S', 2127-2128, 'S', 2129-2130, 'S', 2131-2132, 'S', 2133-2134, 'S', 2135-2136, 'S', 2137-2138, 'S', 2139-2140, 'S', 2141-2142, 'S', 2143-2144, 'S', 2145-2146, 'S', 2147-2148, 'S', 2149-2150, 'S', 2151-2152, 'S', 2153-2154, 'S', 2155-2156, 'S', 2157-2158, 'S', 2159-2160, 'S', 2161-2162, 'S', 2163-2164, 'S', 2165-2166, 'S', 2167-2168, 'S', 2169-2170, 'S', 2171-2172, 'S', 2173-2174, 'S', 2175-2176, 'S', 2177-2178, 'S', 2179-2180, 'S', 2181-2182, 'S', 2183-2184, 'S', 2185-2186, 'S', 2187-2188, 'S', 2189-2190, 'S', 2191-2192, 'S', 2193-2194, 'S', 2195-2196, 'S', 2197-2198, 'S', 2199-2200, 'S', 2201-2202, 'S', 2203-2204, 'S', 2205-2206, 'S', 2207-2208, 'S', 2209-2210, 'S', 2211-2212, 'S', 2213-2214, 'S', 2215-2216, 'S', 2217-2218, 'S', 2219-2220, 'S', 2221-2222, 'S', 2223-2224, 'S', 2225-2226, 'S', 2227-2228, 'S', 2229-2230, 'S', 2231-2232, 'S', 2233-2234, 'S', 2235-2236, 'S', 2237-2238, 'S', 2239-2240, 'S', 2241-2242, 'S', 2243-2244, 'S', 2245-2246, 'S', 2247-2248, 'S', 2249-2250, 'S', 2251-2252, 'S', 2253-2254, 'S', 2255-2256, 'S', 2257-2258, 'S', 2259-2260, 'S', 2261-2262, 'S', 2263-2264, 'S', 2265-2266, 'S', 2267-2268, 'S', 2269-2270, 'S', 2271-2272, 'S', 2273-2274, 'S', 2275-2276, 'S', 2277-2278, 'S', 2279-2280, 'S', 2281-2282, 'S', 2283-2284, 'S', 2285-2286, 'S', 2287-2288, 'S', 2289-2290, 'S', 2291-2292, 'S', 2293-2294, 'S', 2295-2296, 'S', 2297-2298, 'S', 2299-2300, 'S', 2301-2302, 'S', 2303-2304, 'S', 2305-2306, 'S', 2307-2308, 'S', 2309-2310, 'S', 2311-2312, 'S', 2313-2314, 'S', 2315-2316, 'S', 2317-2318, 'S', 2319-2320, 'S', 2321-2322, 'S', 2323-2324, 'S', 2325-2326, 'S', 2327-2328, 'S', 2329-2330, 'S', 2331-2332, 'S', 2333-2334, 'S', 2335-2336, 'S', 2337-2338, 'S', 2339-2340, 'S', 2341-2342, 'S', 2343-2344, 'S', 2345-2346, 'S', 2347-2348, 'S', 2349-2350, 'S', 2351-2352, 'S', 2353-2354, 'S', 2355-2356, 'S', 2357-2358, 'S', 2359-2360, 'S', 2361-2362, 'S', 2363-2364, 'S', 2365-2366, 'S', 2367-2368, 'S', 2369-2370, 'S', 2371-2372, 'S', 2373-2374, 'S', 2375-2376, 'S', 2377-2378, 'S', 2379-2380, 'S', 2381-2382, 'S', 2383-2384, 'S', 2385-2386, 'S', 2387-2388, 'S', 2389-2390, 'S', 2391-2392, 'S', 2393-2394, 'S', 2395-2396, 'S', 2397-2398, 'S', 2399-2400, 'S', 2401-2402, 'S', 2403-2404, 'S', 2405-2406, 'S', 2407-2408, 'S', 2409-2410, 'S', 2411-2412, 'S', 2413-2414, 'S', 2415-2416, 'S', 2417-2418, 'S', 2419-2420, 'S', 2421-2422, 'S', 2423-2424, 'S', 2425-2426, 'S', 2427-2428, 'S', 2429-2430, 'S', 2431-2432, 'S', 2433-2434, 'S', 2435-2436, 'S', 2437-2438, 'S', 2439-2440, 'S', 2441-2442, 'S', 2443-2444, 'S', 2445-2446, 'S', 2447-2448, 'S', 2449-2450, 'S', 2451-2452, 'S', 2453-2454, 'S', 2455-2456, 'S', 2457-2458, 'S', 2459-2460, 'S', 2461-2462, 'S', 2463-2464, 'S', 2465-2466, 'S', 2467-2468, 'S', 2469-2470, 'S', 2471-2472, 'S', 2473-2474, 'S', 2475-2476, 'S', 2477-2478, 'S', 2479-2480, 'S', 2481-2482, 'S', 2483-2484, 'S', 2485-2486, 'S', 2487-2488, 'S', 2489-2490, 'S', 2491-2492, 'S', 2493-2494, 'S', 2495-2496, 'S', 2497-2498, 'S', 2499-2500, 'S', 2501-2502, 'S', 2503-2504, 'S', 2505-2506, 'S', 2507-2508, 'S', 2509-2510, 'S', 2511-2512, 'S', 2513-2514, 'S', 2515-2516, 'S', 2517-2518, 'S', 2519-2520, 'S', 2521-2522, 'S', 2523-2524, 'S', 2525-2526, 'S', 2527-2528, 'S', 2529-2530, 'S', 2531-2532, 'S', 2533-2534, 'S', 2535-2536, 'S', 2537-2538, 'S', 2539-2540, 'S', 2541-2542, 'S', 2543-2544, 'S', 2545-2546, 'S', 2547-2548, 'S', 2549-2550, 'S', 2551-2552, 'S', 2553-2554, 'S', 2555-2556, 'S', 2557-2558, 'S', 2559-2560, 'S', 2561-2562, 'S', 2563-2564, 'S', 2565-2566, 'S', 2567-2568, 'S', 2569-2570, 'S', 2571-2572, 'S', 2573-2574, 'S', 2575-2576, 'S', 2577-2578, 'S', 2579-2580, 'S', 2581-2582, 'S', 2583-2584, 'S', 2585-2586, 'S', 2587-2588, 'S', 2589-2590, 'S', 2591-2592, 'S', 2593-2594, 'S', 2595-2596, 'S', 2597-2598, 'S', 2599-2600, 'S', 2601-2602, 'S', 2603-2604, 'S', 2605-2606, 'S', 2607-2608, 'S', 2609-2610, 'S', 2611-2612, 'S', 2613-2614, 'S', 2615-2616, 'S', 2617-2618, 'S', 2619-2620, 'S', 2621-2622, 'S', 2623-2624, 'S', 2625-2626, 'S', 2627-2628, 'S', 2629-2630, 'S', 2631-2632, 'S', 2633-2634, 'S', 2635-2636, 'S', 2637-2638, 'S', 2639-2640, 'S', 2641-2642, 'S', 2643-2644, 'S', 2645-2646, 'S', 2647-2648, 'S', 2649-2650, 'S', 2651-2652, 'S', 2653-2654, 'S', 2655-2656, 'S', 2657-2658, 'S', 2659-2660, 'S', 2661-2662, 'S', 2663-2664, 'S', 2665-2666, 'S', 2667-2668, 'S', 2669-2670, 'S', 2671-2672, 'S', 2673-2674, 'S', 2675-2676, 'S', 2677-2678, 'S', 2679-2680, 'S', 2681-2682, 'S', 2683-2684, 'S', 2685-2686, 'S', 2687-2688, 'S', 2689-2690, 'S', 2691-2692, 'S', 2693-2694, 'S', 2695-2696, 'S', 2697-2698, 'S', 2699-2700, 'S', 2701-2702, 'S', 2703-2704, 'S', 2705-2706, 'S', 2707-2708, 'S', 2709-2710, 'S', 2711-2712, 'S', 2713-2714, 'S', 2715-2716, 'S', 2717-2718, 'S', 2719-2720, 'S', 2721-2722, 'S', 2723-2724, 'S', 2725-2726, 'S', 2727-2728, 'S', 2729-2730, 'S', 2731-2732, 'S', 2733-2734, 'S', 2735-2736, 'S', 2737-2738, 'S', 2739-2740, 'S', 2741-2742, 'S', 2743-2744, 'S', 2745-2746, 'S', 2747-2748, 'S', 2749-2750, 'S', 2751-2752, 'S', 2753-2754, 'S', 2755-2756, 'S', 2757-2758, 'S', 2759-2760, 'S', 2761-2762, 'S', 2763-2764, 'S', 2765-2766, 'S', 2767-2768, 'S', 2769-2770, 'S', 2771-2772, 'S', 2773-2774, 'S', 2775-2776, 'S', 2777-2778, 'S', 2779-2780, 'S', 2781-2782, 'S', 2783-2784, 'S', 2785-2786, 'S', 2787-2788, 'S', 2789-2790, 'S', 2791-2792, 'S', 2793-2794, 'S', 2795-2796, 'S', 2797-2798, 'S', 2799-2800, 'S', 2801-2802, 'S',

THIS PAGE LEFT BLANK

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:13:00 ; Search time 134.566 Seconds
(without alignments)
1295.016 Million cell updates/sec

Title: US-10-677-877A-4

Perfect score: 1329

Sequence: 1 RSDANVRDLEVDITLKS.....LDVGAPDQEGFDVGPVCF 247

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1317	99.1	1461	2	O76045 HUMAN
2	1317	99.1	1464	1	CO1A1_HUMAN
3	1313	98.8	1467	2	O59F64 HUMAN
4	1295	97.4	1464	2	O8N473 HUMAN
5	1290	97.1	287	2	O95ND8 BOVIN
6	1290	97.1	1460	1	CO1A1_CANFA
7	1238	93.2	1453	2	O63079 RAT
8	1235	92.9	589	2	O99LL6_MOUSE
9	1235	92.9	1225	2	O6PCL3_MOUSE
10	1235	92.9	1453	2	O81079_MOUSE
11	1225	92.2	1453	1	CO1A1_CHICK
12	1218	91.6	1453	1	CO1A1_MOUSE
13	1205	90.7	325	2	O60785_MOUSE
14	1205	90.7	1450	2	O9Y1B4_CYNPY
15	1185	89.2	1449	2	O640B2_XENTR
16	1177	88.6	1445	2	O93251_RANCA
17	1159	87.2	1449	2	O802B5_XENLA
18	1153	86.8	809	2	O93485_ONCMY
19	1153	86.8	1449	2	O910C0_ONCMY
20	1151	86.6	1447	2	O91B91_XENLA
21	1146	86.2	481	2	O90YJ1_BRARE
22	1146	86.2	1447	2	O6P4U1_BRARE
23	1146	86.2	1447	2	O6ULJ5_BRARE
24	1145	86.2	1399	2	O4SB89_TETNG
25	1130	85.0	1447	2	O5NT96_PAROL
26	1078.5	81.2	1463	2	O4W6W6_9CHON
27	1063	80.0	678	2	O93486_ONCMY
28	1063	80.0	1458	2	O910B9_ONCMY
29	1058	79.6	441	2	O90Y19_BRARE
30	1058	79.6	1449	2	O6NZ15_BRARE
31	1058	79.6	1449	2	O6PE19_BRARE

32	1048	78.9	1545	2	O4RYE7_TETNG
33	1006.5	75.7	369	1	CO2A1_CHICK
34	1005.5	75.7	1269	2	O7T2Z7_CHICK
35	993.5	74.8	1420	2	O90W37_GALLU
36	990.5	74.5	1487	2	O77753_CANFA
37	984.5	74.1	1418	2	O9W7R9_CYNPY
38	982.5	73.9	268	2	O96IT5_HUMAN
39	982.5	73.9	1418	1	CO2A1_HUMAN
40	982.5	73.9	1487	2	Q14047_HUMAN
41	979.5	73.7	1418	2	Q28396_EQUUS
42	973.5	73.3	1419	2	O63123_RAT
43	967.5	72.8	826	2	O8K0N6_MOUSE
44	967.5	72.8	886	2	O8CRF7_MOUSE
45	967.5	72.8	1419	2	O80VY3_MOUSE

ALIGNMENTS

RESULT 1
O76045_HUMAN
ID O76045_HUMAN PRELIMINARY; PRT; 1461 AA.
AC O76045;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Pro alpha 1(I) collagen.
GN Name=COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
Promoter structure, Alu repeats, and polymorphic transcripts.";
J. Biol. Chem. 260:2315-2320(1985).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
twenty-five exons of the human pro alpha 1(I) collagen gene
(COL1A1).";
Gene 67:105-115(1988).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
of human type I procollagen.";
Biochem. J. 253:919-922(1988).
RN [4]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
COL1A1 gene bind cell-specific nuclear proteins.";
FEBS Lett. 279:9-13(1991).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=92157916; PubMed=1787829;
RA Westerhausen A., Constantinou C.D., Pack M., Peng M.Z., Hanning C.,
Olson A.S., Prockop D.J.;
RT "Completion of the last half of the structure of the human gene for
the Pro alpha 1 (I) chain of type I procollagen (COL1A1).";
Matrix 11:375-379(1991).
RN [6]

```
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98107942; PubMed=9443882;
RA Korkko J., Ala-Kokko L., De Paep A., Nuytink L., Earley J.,
RA Prockop D.J.;
RT "Analysis of the COL1A1 and COL1A2 genes by PCR amplification and
RT scanning by conformation-sensitive gel electrophoresis identifies only
RT COL1A1 mutations in 15 patients with osteogenesis imperfecta type I:
RT identification of common sequences of null-allele mutations.";
RL Am. J. Hum. Genet. 62:98-110(1998).
RN [7]
RP NUCLEOTIDE SEQUENCE.
RA Korkko J.M., Earley J.J., Nuytink L., DePaep A., Prockop D.J.,
RA Ala-Kokko L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RL ENBL; AF017178; AB94054.2; -; Genomic_DNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VMC; 1.
DR PROSITE; PS01208; VWF 1; 1.
DR PROSITE; PS0184; VWF 2; 1.
KW Collagen.
SQ SEQUENCE 1461 AA; 138629 MW; 9ACF6DE30EA78E21 CRC64;

Query Match 99.1%; Score 1317; DB 2; Length 1461;
Best Local Similarity 99.6%; Pred. No. 5.1e-111;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVDRDLVDVTTKLSLQOINIRSPGSRKNPARTCRDLKMSDWKSGEYWDPN 62
Db 1217 DANVDRDLVDVTTKLSLQOINIRSPGSRKNPARTCRDLKMSDWKSGEYWDPN 1276

QY 63 QGCLNDAIKVFCNMEGETCVYTPQPSVAQKNWYISKPKDHRVWFGESMTDGFQFEYG 122
Db 1277 QGCLNDAIKVFCNMEGETCVYTPQPSVAQKNWYISKPKDHRVWFGESMTDGFQFEYG 1336

QY 123 QGSDPADVAIQLTFLRLMSTEAQNITYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
Db 1337 QGSDPADVAIQLTFLRLMSTEAQNITYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1396

QY 183 AEGNSRFTYVTDGCTSHTGANGKTVIEYTKTSRLPIIDVAPLDVGAPDQGFQDVG 242
Db 1397 AEGNSRFTYVTDGCTSHTGANGKTVIEYTKTSRLPIIDVAPLDVGAPDQGFQDVG 1456

QY 243 PVCFL 247
Db 1457 PVCFL 1461

RESULT 2
COL1A1 HUMAN STANDARD; PRT: 1464 AA.
AC P02452; F78441; Q13896; Q13902; Q13903; Q14037; Q14992; Q15176;
AC Q15201; Q16050; Q7KZ30; Q7KZ34; Q81V15; Q9UML6; Q9UMW7;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name:COL1A1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RP NUCLEOTIDE SEQUENCE, AND VARIANT SER-1434.
RA Dalgleish R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE OF 1-589.
RX MEDLINE=88329734; PubMed=2843432; DOI=10.1016/0378-1119(88)90013-3;
RA D'Alessio M., Bernard M.P., Pretorius P.J., de Wet W., Ramirez F.,
RA Pretorius P.J.;
RT "Complete nucleotide sequence of the region encompassing the first
RT twenty-five exons of the human pro alpha 1(I) collagen gene
RT (COL1A1).";
RL Gene 67:105-115(1988).
RN [3]
RP NUCLEOTIDE SEQUENCE OF 1-472.
RX MEDLINE=89025644; PubMed=3178743;
RA Tromp G., Kuivaniemi H., Stacey A., Shikata H., Baldwin C.T.,
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I) chain
RT of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1-181.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [5]
RP PROTEIN SEQUENCE OF 162-301, AND PYRROLIDONE CARBOXYLIC ACID.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA Click E.M., Bornstein P.;
RT "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [6]
RP NUCLEOTIDE SEQUENCE OF 425-1464.
RX MEDLINE=84080385; PubMed=6689127;
RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA Prockop D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT of structures that are conserved during evolution.";
RL Biochemistry 22:5213-5223(1983).
RN [7]
RP NUCLEOTIDE SEQUENCE OF 472-607.
RX MEDLINE=85104934; PubMed=2981843;
RA Chu M.-L., Gargiulo V., Williams C.J., Ramirez F.;
RT "Multixon deletion in an osteogenesis imperfecta variant with
RT increased type III collagen mRNA.";
RL J. Biol. Chem. 260:691-694(1985).
RN [8]
RP NUCLEOTIDE SEQUENCE OF 488-625.
RX MEDLINE=85190598; PubMed=3857621;
RA Barsh G.S., Roush C.L., Bonadio J., Byers P.H., Gelinas R.E.;
RT "Intron-mediated recombination may cause a deletion in an alpha 1 type
RT I collagen chain in a lethal form of osteogenesis imperfecta.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2870-2874(1985).
RN [9]
RP NUCLEOTIDE SEQUENCE OF 1179-1464, VARIANTS OI-II HIS-1277; ARG-1388
RX AND 1337-GLU-TYR-1338 DEL, AND VARIANT SER-1434.
RX MEDLINE=93352646; PubMed=8349697;
RA Chessler S.D., Wallis G.A., Byers P.H.;
RT "Mutations in the carboxyl-terminal propeptide of the pro alpha 1(I)
RT chain of type I collagen result in defective chain association and
RT produce lethal osteogenesis imperfecta.";
RL J. Biol. Chem. 268:18218-18225(1993).
RN [10]
RP NUCLEOTIDE SEQUENCE OF 1229-1454.
RC TISSUE=Bone;
RX
```

RA MEDLINE=88124208; PubMed=3340531;
RA Maekelae J.K., Raasina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349 (1988).
RN [11]
RP NUCLEOTIDE SEQUENCE OF 1-34.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, AluI repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320 (1985).
RN [12]
RP NUCLEOTIDE SEQUENCE OF 1-34.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devarayalu S., Gelinas R.E.;
RT "Regulatory elements in the first intron contribute to transcriptional
RT control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873 (1987).
RN [13]
RP NUCLEOTIDE SEQUENCE OF 1-44.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157 (1987).
RN [14]
RP PROTEIN SEQUENCE OF 33-52.
RX MEDLINE=90202908; PubMed=2318855;
RA Wirtz M.K., Keene D.R., Hori H., Glanville R.W., Steinmann B.,
RA Rao V.H., Hollister D.W.;
RT "In vivo and in vitro noncovalent association of excised alpha 1 (I)
RT amino-terminal propeptides with mutant pN alpha 2(I) collagen chains
RT in native mutant collagen in a case of Ehlers-Danlos syndrome, type
RT VII.";
RL J. Biol. Chem. 265:6312-6317 (1990).
RN [15]
RP NUCLEOTIDE SEQUENCE OF 156-183.
RX MEDLINE=89356643; PubMed=2767050;
RA Weil D., D'Alessio M., Ramirez F., de Wet W., Cole W.G., Chan D.,
RA Bateman J.F.;
RT "A base substitution in the exon of a collagen gene causes alternative
RT splicing and generates a structurally abnormal polypeptide in a
RT patient with Ehlers-Danlos syndrome type VII.";
RL EMBO J. 8:1705-1710 (1989).
RN [16]
RP PROTEIN SEQUENCE OF 175-187 AND 274-289.
RX MEDLINE=9032436; PubMed=2169412;
RA Baetge B., Notbohm H., Diebold J., Lehmann H., Bodo M., Deutzmann R.,
RA Muller P.K.;
RT "A critical crosslink region in human-bone-derived collagen type I.
RT Specific cleavage site at residue Leu95.";
RL Eur. J. Biochem. 192:153-159 (1990).
RN [17]
RP PROTEIN SEQUENCE OF 263-268.
RC TISSUE=Skin;
RX MEDLINE=71001508; PubMed=4319110;
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT "A comparative study of glycopeptides derived from selected vertebrate
RT collagens. A possible role of the carbohydrate in fibril formation.";
RL J. Biol. Chem. 245:5042-5048 (1970).
RN [18]
RP NUCLEOTIDE SEQUENCE OF 281-302; 402-420; 823-842; 924-944; 1026-1045
AND 1143-1162.
RX MEDLINE=90326017; PubMed=2374517;
RA Labhard M.E., Hollister D.W.;
RT "Segmental amplification of the entire helical and telopeptide regions
RT of the cDNA for human alpha 1 (I) collagen.";
RL Matrix 10:124-130 (1990).
RN [19]
RP NUCLEOTIDE SEQUENCE OF 425-430; 965-1024; 999-1039 AND 1453-1464.
RX MEDLINE=83064528; PubMed=6183642;

RA Chu M.-L., Myers J.C., Bernard M.P., Ding J.-F., Ramirez F.;
RT "Cloning and characterization of five overlapping cDNAs specific for
RT the human pro alpha 1(I) collagen chain.";
RL Nucleic Acids Res. 10:5925-5934 (1982).
RN [20]
RP NUCLEOTIDE SEQUENCE OF 710-745, AND VARIANT OI-II ARG-728.
RX MEDLINE=90252792; PubMed=2339700;
RA Wallis G.A., Starman B.J., Zinn A.B., Byers P.H.;
RT "Variable expression of osteogenesis imperfecta in a nuclear family is
RT explained by somatic mosaicism for a lethal point mutation in the
RT alpha 1(I) gene (COL1A1) of type I collagen in a parent.";
RL Am. J. Hum. Genet. 46:1034-1040 (1990).
RN [21]
RP NUCLEOTIDE SEQUENCE OF 746-781, AND VARIANT OI-III SER-767.
RX MEDLINE=95187161; PubMed=7881420;
RA Forlino A., Zolezzi F., Valli M., Pignatti P.F., Cetta G.,
RA Brunelli P.C., Mottes M.;
RT "Severe (type III) osteogenesis imperfecta due to glycine
RT substitutions in the central domain of the collagen triple helix.";
RL Hum. Mol. Genet. 3:2201-2206 (1994).
RN [22]
RP NUCLEOTIDE SEQUENCE OF 1187-1220, AND VARIANT CYS-1195.
RX MEDLINE=89008319; PubMed=3170557;
RA Cohn D.H., Apone S., Byre D.R., Starman B.J., Andreassen P.,
RA Charbonneau H., Nicholls A.C., Pope F.M., Byers P.H.;
RT "Substitution of cysteine for glycine within the carboxyl-terminal
RT telopeptide of the alpha 1 chain of type I collagen produces mild
RT osteogenesis imperfecta.";
RL J. Biol. Chem. 263:14605-14607 (1988).
RN [23]
RP NUCLEOTIDE SEQUENCE OF 1440-1464.
RX MEDLINE=90110490; PubMed=2295701;
RA Willing M.C., Cohn D.H., Byers P.H.;
RT "Frameshift mutation near the 3' end of the COL1A1 gene of type I
RT collagen predicts an elongated pro alpha 1(I) chain and results in
RT osteogenesis imperfecta type I.";
RL J. Clin. Invest. 85:282-290 (1990).
RN [24]
RP NUCLEOTIDE SEQUENCE OF 1454-1464.
RX MEDLINE=91138770; PubMed=1995349; DOI=10.1016/0014-5793(91)80237-W;
RA Maatta A., Bornstein P., Penttinen R.P.;
RT "Highly conserved sequences in the 3'-untranslated region of the
Query Match 99.1%; Score 1317; DB 1; Length 1464;
Best Local Similarity 99.6%; Pred. No. 5.1e-111;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 3 DANVVRDRDLVDVTLKSLSQIENIRSPGSRKNPARTCRDLKMSHSDKSGEYWDPN 62
Db 1220 DANVVRDRDLVDVTLKSLSQIENIRSPGSRKNPARTCRDLKMSHSDKSGEYWDPN 1279
Qy 63 QGCNLDIAIKVFCNMETGETCVYTPQPSVAQKNWYISKPNKDKRHWVFGESMTDGFQFEYG 122
Db 1280 QGCNLDIAIKVFCNMETGETCVYTPQPSVAQKNWYISKPNKDKRHWVFGESMTDGFQFEYG 1339
Qy 123 GQSGDPADVAIQLTFTLRMLSTEAQNITYHCNKSVAIYMDQQTGNLKKALLKGSNEIEIR 182
Db 1340 GQSGDPADVAIQLTFTLRMLSTEAQNITYHCNKSVAIYMDQQTGNLKKALLKGSNEIEIR 1399
Qy 183 AEGNSBFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGADPQEFQFVVG 242
Db 1400 AEGNSBFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGADPQEFQFVVG 1459
Qy 243 PVCFL 247
Db 1460 PVCFL 1464
RESULT 3
Q59F64 HUMAN
ID Q59F64 HUMAN PRELIMINARY; PRT; 1467 AA.
AC Q59F64;
DT 10-MAY-2005 (TREMBLrel. 30, Created)


```

OC Pecora; Bovidae; Bovinae; Bos.
OK NCBI_TaxID=9913;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Grosse-Hovest L., Brem G.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ312109; CAC38832.1; -; Genomic_DNA.
DR EMBL; AJ312110; CAC38832.1; JOINED; Genomic_DNA.
DR EMBL; AJ312111; CAC38832.1; JOINED; Genomic_DNA.
DR EMBL; AJ312112; CAC38832.1; JOINED; Genomic_DNA.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008160; Collagen.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI_1.
KW Collagen.
FT NON_TER
SQ SEQUENCE 287 AA; 31698 MW; A9122CAEB7DC3DAE CRC64;

Query Match 97.1%; Score 1290; DB 2; Length 287;
Best Local Similarity 97.1%; Pred. No. 1.8e-109;
Matches 238; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DANVVRDRDLEVDVTLKSLSQIENIRSPGSRKNPARTCRDLKMSHDSKSGEYWDPN 62
Db 43 DANVVRDRDLEVDVTLKSLSQIENIRSPGSRKNPARTCRDLKMSHDSKSGEYWDPN 102

Qy 63 QGCMNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQPEYG 122
Db 103 QGCMNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQPEYG 162

Qy 123 QGQSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 182
Db 163 QGQSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 222

Qy 183 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 223 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 282

Qy 243 PFCFL 247
Db 283 PFCFL 287

RESULT 6
COL1A1_CANFA STANDARD; PRT; 1460 AA.
AC Q9XJ7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
OC Canis.
OK NCBI_TaxID=9615;
RN [1]
RP NUCLEOTIDE SEQUENCE, AND VARIANT OI ALA-208.
RC TISSUE=Skin;
RX MEDLINE=21023337; PubMed=11147834; DOI=10.1006/abbi.2000.2099;
RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
RT "Sequence of normal canine COL1A1 cDNA and identification of a
RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case of
RT canine osteogenesis imperfecta."
RL Arch. Biochem. Biophys. 384:37-46(2000).
CC -!- FUNCTION: type I collagen is a member of group I collagen
CC (fibrillar forming collagen).

```

```

CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
CC (OI).
CC -!- SIMILARITY: Belongs to the fibrillar collagen family.
CC -!- SIMILARITY: Contains 1 VWFC domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; AF153062; AAD34619.1; -; mRNA.
CC Ensembl; ENSCARG00000017018; Canis familiaris.
CC InterPro; IPR008161; Clg helix.
CC InterPro; IPR008160; Collagen.
CC InterPro; IPR000885; Fib_collagen_C.
CC InterPro; IPR001007; VWFC_C.
CC Pfam; PF01410; COLFI; 1.
CC ProDom; PD000007; Clg helix; 2.
CC ProSITE; PS01208; VWFC_1; 1.
CC PROSITE; PS0184; VWFC_2; 1.
CC Collagen; Disease mutation; Extracellular matrix; Glycoprotein;
KW Hydroxylation; Pyrrolidone carboxylic acid; Repeat; Signal;
KW Structural protein.
FT SIGNAL 1 22 By similarity.
FT PROPEP 23 157 N-terminal propeptide.
FT CHAIN 158 1214 Collagen alpha 1(I) chain.
FT PROPEP 1215 1460 C-terminal propeptide.
FT DOMAIN 34 92 VWFC.
FT REGION 158 174 Nonhelical region (N-terminal).
FT REGION 175 1188 Triple-helical region.
FT REGION 1189 1214 Nonhelical region (C-terminal).
FT MOTIF 741 743 Cell attachment site (Potential).
FT MOTIF 1089 1091 Cell attachment site (Potential).
FT MOD_RES 158 158 Pyrrolidone carboxylic acid (By
FT similarity).
FT MOD_RES 166 166 Allysine (By similarity).
FT MOD_RES 261 261 5-hydroxylysine (By similarity).
FT MOD_RES 1160 1160 O-linked (Gal..) (By similarity).
FT CARBOHYD 261 261 N-linked (GlcNAc..) (By similarity).
FT VARIANT 208 208 G -> A (in OI; severe).
SQ SEQUENCE 1460 AA; 138762 MW; 58E3674D2B570697 CRC64;

Query Match 97.1%; Score 1290; DB 1; Length 1460;
Best Local Similarity 96.7%; Pred. No. 1.5e-108;
Matches 237; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DANVVRDRDLEVDVTLKSLSQIENIRSPGSRKNPARTCRDLKMSHDSKSGEYWDPN 62
Db 1216 DANVVRDRDLEVDVTLKSLSQIENIRSPGSRKNPARTCRDLKMSHDSKSGEYWDPN 1275

Qy 63 QGCMNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQPEYG 122
Db 1276 QGCMNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQPEYG 1335

Qy 123 QGQSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 182
Db 1336 QGQSDPADVAIQLTFLRLMSTEASQNTYHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 1395

Qy 183 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1396 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 1455

Qy 243 PFCFL 247
Db 1456 PFCFL 1460

```



```

RESULT 7
Q63079 RAT PRELIMINARY; PRT; 1453 AA.
ID Q63079 RAT PRELIMINARY; PRT; 1453 AA.
AC Q63079;
DT 01-JUN-1996 (TrEMBLrel. 01, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Collagen alpha1 (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Sprague-Dawley; TISSUE=Bone, and Tooth;
RX MEDLINE=99163824; PubMed=10065941;
RA Brandsten C., Lundmark C., Christerson C., Hammarstrom L., Wurtz T.;
RT "Expression of collagen alpha1(I) mRNA variants during tooth and bone
formation in the rat."
RL J. Dent. Res. 78:11-19(1999).
DR EMBL; Z78279; CAB01633.1; -; mRNA.
DR Ensembl; ENSRNOG0000003897; Rattus norvegicus.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0005515; F:protein binding; IEA.
DR GO; GO:0007155; P:cell adhesion; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR InterPro; IPR001007; VF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg helix; 3.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; WVC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 1453 AA; 137886 MW; E6896BDC19A4A1D8 CRC64;

Query Match 93.2%; Score 1238; DB 2; Length 1453;
Best Local Similarity 91.0%; Pred. No. 8.5e-104;
Matches 223; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Qy 3 DANVVRDRDLEVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db 1209 DANVVRDRDLEVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1268

Qy 63 QGCNLDIAKVCNMTGETCVYPTOPSAQKNWYISKNPDKRHVWFGESMTDGFQFEYG 122
Db 1269 QGCNLDIAKVCNMTGETCVYPTOPSAQKNWYISKNPDKRHVWFGESMTDGFQFEYG 1328

Qy 123 GQGSDDPADVAIQITFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNIEIR 182
Db 1329 SEGSDPADVAIQITFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNIEIR 1388

Qy 183 AEGNSRFTSYVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPODFGFDVG 242
Db 1389 GEGNSRFTSYTLVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPODFGFDVG 1448

Qy 243 PQCFL 247
Db 1449 PACFV 1453

RESULT 8
Q99LL6_MOUSE
Q99LL6_MOUSE PRELIMINARY; PRT; 589 AA.
Q99LL6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Colial protein (Fragment).
GN Name=Colial;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N;
TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
virgin mouse. Taken by biopsy.;
RX MEDLINE=22388957; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Klausner R.L., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Rosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalinski D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FVB/N;
TISSUE=Mammary tumor. Metallothionien-TGF alpha model. 10 month old
virgin mouse. Taken by biopsy.;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003198; AA03198.1; -; mRNA.
DR MGI; MGI:88467; Colial.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD000007; Clg helix; 1.
DR ProDom; PD002078; Fib collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 589 AA; 58805 MW; 81847495E5E05CEF CRC64;

Query Match 92.9%; Score 1235; DB 2; Length 589;
Best Local Similarity 91.0%; Pred. No. 4.9e-104;
Matches 223; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

Qy 3 DANVVRDRDLEVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db 345 DANVVRDRDLEVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 404

Qy 63 QGCNLDIAKVCNMTGETCVYPTOPSAQKNWYISKNPDKRHVWFGESMTDGFQFEYG 122
Db 405 QGCNLDIAKVCNMTGETCVYPTOPSAQKNWYISKNPDKRHVWFGESMTDGFQFEYG 464

Qy 123 GQGSDDPADVAIQITFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNIEIR 182

```

Db	465	SEGSDFADVAIQTLFURLMSTSEASONITTHCKNSVAYNDQDTGNLKALLQGSNEI	ELR	524		
Qy	183	ACGNSRFTTTSVTVGCTSHGAGTKVIEYKTKSSRLPIIDVAPLDVGADQDFGFDVG	242			
Db	525	CGNSRFTTSTLVGCTSHGAGTKVIEYKTKSSRLPIIDVAPLDVGADQDFGFDVG	584			
Qy	243	PVCFL 247				
Db	585	PACFV 589				
RESULT 9						
Q6PCL3	MOUSE					
ID	Q6PCL3	MOUSE PRELIMINARY;	PRT;	1225 AA.		
AC	Q6PCL3;					
DT	05-JUL-2004	(T-EMBLrel. 27, Created)				
DT	05-JUL-2004	(T-EMBLrel. 27, Last sequence update)				
DT	05-JUL-2004	(T-EMBLrel. 27, Last annotation update)				
DE	Collal protein.					
GN	Name=Collal;					
OS	Mus musculus (Mouse).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;					
OC	Muridae; Murinae; Mus.					
OX	NCBI_TaxID=10090;					
RN	[1]					
RP	NUCLEOTIDE SEQUENCE.					
RC	STRAIN=Mix FVB/N;					
RC	TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;					
RK	MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.2426038999;					
RA	Straussberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,					
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,					
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,					
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,					
RA	Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,					
RA	Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,					
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,					
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,					
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,					
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,					
RA	Villalón D.K., Murny D.M., Sodergren E.J., Lu X., Gibbs R.A.,					
RA	Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,					
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,					
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,					
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,					
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,					
RA	Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;					
RT	"Generation and initial analysis of more than 15,000 full-length human					
RT	and mouse cDNA sequences."					
RN	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).					
RN	[2]					
RP	NUCLEOTIDE SEQUENCE.					
RC	STRAIN=Mix FVB/N;					
RC	TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;					
RA	Straussberg R.;					
DR	Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; BC059281; AAH59281.1; -, mRNA.					
DR	MGI; MGI:88467; Collal.					
DR	GO; GO:0005615; C:extracellular space; TAS.					
DR	InterPro; IPR008161; Clg helix.					
DR	InterPro; IPR008160; Collagen.					
DR	InterPro; IPR000885; Fib collagen_C.					
DR	InterPro; IPR001007; VF_C.					
DR	Pfam; PF01410; COLFI; 1.					
DR	Pfam; PF01391; Collagen; 14.					
DR	Pfam; PF00093; VWC; 1.					
DR	ProDom; PD000007; Clg helix; 2.					
DR	ProDom; PD002078; Fib collagen_C;					

```
RA Chapman J.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RA Smith M.;
RL Submitted (FEB-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC050014; AAH50014.1; -; mRNA.
DR EMBL; AL606480; CA123970.1; -; Genomic DNA.
DR EMBL; AL662790; CA125880.1; -; Genomic DNA.
DR EMBL; AL662790; CA123970.1; JOINED; Genomic DNA.
DR EMBL; AL606480; CA125880.1; JOINED; Genomic DNA.
DR EMBL; ENSMUSG0000001506; Mus musculus.
DR MGI; MGI:88467; Coll1a1.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS01084; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1453 AA; 138032 MW; 0B7F06BB9A1D5EA CRC64;
Query Match 92.9%; Score 1235; DB 2; Length 1453;
Best Local Similarity 91.0%; Pred. No. 1.6e-103;
Matches 223; Conservative 12; Mismatches 10; Indels 0; Gaps 0;
Qy 3 DANVVRDRLEVDTTLSLQSIQINIRSPGSRKNPARTCRDLKWCCHSDWKSGEYWDPN 62
Db 1209 DANVVRDRLEVDTTLSLQSIQINIRSPGSRKNPARTCRDLKWCCHSDWKSGEYWDPN 1268
Qy 63 QGCNLDIAKVPCHMETGETCVYPTOPSAQKNYISKPKDCKHVVFGESMTGDFPEYG 122
Db 1269 QGCNLDIAKVPCHMETGETCVYPTOPSAQKNYISKPKDCKHVVFGESMTGDFPEYG 1328
Qy 123 GCGSDPADVAIQTLFLRMTEASQNTYTHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1329 SEGSDPADVAIQTLFLRMTEASQNTYTHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1388
Qy 183 AEGNSRFTSYVTVDGCTSHTGKAGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEPGFDVG 242
Db 1389 GEGNSRFTSYVTVDGCTSHTGKAGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEPGFDVG 1448
Qy 243 PVCFEL 247
Db 1449 PACFV 1453
RESULT 11
COL1A1_CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN Name=COL1A1;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP NUCLEOTIDE SEQUENCE OF 1-153.
RX MEDLINE=88056316; PubMed=3678834; DOI=10.1016/0378-1119(87)90159-4;
RA Finer M.H., Boedtker H., Doty P.;
"Construction and characterization of cDNA clones encoding the 5' end of the chicken pro alpha 1(I) collagen mRNA.";
Gene 56:71-78(1987).
[2]
RN NUCLEOTIDE SEQUENCE OF 1-144.
RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
"Unusual DNA sequences located within the promoter region and the first intron of the chicken pro-alpha 1(I) collagen gene.";
J. Biol. Chem. 262:13323-13332(1987).
[3]
RN PROTEIN SEQUENCE OF 152-1187.
RX MEDLINE=82231995; PubMed=7093229;
RA Higberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M., Kang A.H., Gross J.;
"Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the complete primary structure of the helical portion of the chick skin collagen alpha 1(I) chain.";
Biochemistry 21:2048-2055(1982).
[4]
RN PROTEIN SEQUENCE OF 1200-1205.
RX MEDLINE=72243016; PubMed=5047697;
RA Eyre D.R., Glimcher M.J.;
"Evidence for a previously undetected sequence at the carboxyterminus of the alpha 1 chain of chicken bone collagen.";
Biochem. Biophys. Res. Commun. 48:720-726(1972).
[5]
RN NUCLEOTIDE SEQUENCE OF 981-1453.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
"Sequence determination and analysis of the 3' region of chicken pro-alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids including the carboxy-terminal propeptide sequences.";
Biochemistry 20:996-1006(1981).
[6]
RN NUCLEOTIDE SEQUENCE OF 1311-1453.
RX MEDLINE=80134546; PubMed=6987088; DOI=10.1016/0014-5793(80)80761-7;
RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T., Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
"Nucleotide sequence of a collagen cDNA-fragment coding for the carboxyl end of pro alpha 1(I)-chains.";
FEBS Lett. 111:61-65(1980).
CC -I- FUNCTION: Type I collagen is a member of group I collagen (fibrillar forming collagen).
CC -I- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -I- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and bones. In bones the fibrils are mineralized with calcium hydroxyapatite.
CC -I- PTM: Proline residues at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains. Pro-1153 is the only 3-hydroxypro and the only of the hydroxylated proline in position X.
CC -I- SIMILARITY: Belongs to the fibrillar collagen family.
CC -I- SIMILARITY: Contains 1 VWF domain.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.
-----
DR EMBL; M17839; AAA48704.1; -; Genomic DNA.
DR EMBL; M17839; AAA48704.1; JOINED; Genomic DNA.
DR EMBL; V00401; CA423695.1; -; mRNA.
DR EMBL; M10571; AAA48671.1; ALT_SEQ; mRNA.
DR EMBL; M17607; AAA48672.1; -; mRNA.
DR PIR; A27179; A27179.
DR PIR; I50629; I50629.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
```

DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF00093; WVC; 1.
 DR ProDom; ED000007; Clg_helix; 2.
 DR ProDom; ED002078; Fib_Collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; WVC; 1.
 DR PROSITE; PS01208; WVC; 1.
 DR PROSITE; PS0184; WVC; 2.
 KW Collagen; Direct protein sequencing; Extracellular matrix;
 KW Glycoprotein; Hydroxylation; Pyridone carboxylic acid; Repeat;
 KW Signal; Structural protein.
 FT SIGNAL 1 22 N-terminal propeptide.
 FT PROPEP 23 151 Collagen alpha 1(I) chain.
 FT CHAIN 152 1205 C-terminal propeptide.
 FT PROPEP 1206 1453 WVC.
 FT DOMAIN 31 89 Pyridone carboxylic acid.
 FT MOD_RES 152 152 Allylsine (By similarity).
 FT MOD_RES 160 160 5-hydroxylysine (By similarity).
 FT MOD_RES 254 254 5-hydroxylysine (Potential).
 FT MOD_RES 851 851 5-hydroxylysine (Potential).
 FT MOD_RES 1081 1081 Hydroxyproline (Potential).
 FT MOD_RES 1097 1097 3-hydroxyproline.
 FT MOD_RES 1153 1153 O-linked (Gal...) (By similarity).
 FT CARBOHYD 254 254 N-linked (GlcNAc...) (By similarity).
 FT CARBOHYD 1354 1354 F -> L (in Ref. 5).
 FT CONFLICT 1187 1187 Q -> H (in Ref. 6).
 FT CONFLICT 1441 1441
 SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;
 Query Match 92.2%; Score 1225; DB 1; Length 1453;
 Best Local Similarity 89.0%; Pred. No. 1.3e-102;
 Matches 216; Conservative 20; Mismatches 7; Indels 0; Gaps 0;
 Qy 3 DANYVRDLEVDITLKSLSQIENIRSPGSRKNPARTCRDLKMSDKSGEYWDPN 62
 Db 1209 DANYVRDLEVDITLKSLSQIENIRSPGSRKNPARTCRDLKMSDKSGEYWDPN 1268
 Qy 63 QGCMNDAIKVFCNNMETGETCYPTQPSVAQKNWYISKNPKDKRHWFGESMTGDFQFEYG 122
 Db 1269 QGCMNDAIKVFCNNMETGETCYPTQPSVAQKNWYISKNPKDKRHWFGESMTGDFQFEYG 1328
 Qy 123 GQSGDPADVAIQLTFLRLMSTSEASONITYHCKNSVAYMDQOTGNLKKALLKGSNELEIR 182
 Db 1329 GEGSNPDAVAIQLTFLRLMSTSEASONITYHCKNSVAYMDQOTGNLKKALLKGSNELEIR 1388
 Qy 183 AEGNSRTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPDIIDVAPLDVGAPDOEFGFDVG 242
 Db 1389 AEGNSRTYGTEDGCTSHTGANGKTVIEYKTKSSRLPDIIDVAPLDVGAPDOEFGFDVG 1448
 Qy 243 PVCF 247
 Db 1449 PVCF 1453
 RESULT 12
 COL1A1 MOUSE
 ID COL1A1 MOUSE STANDARD; PRT: 1453 AA.
 AC P11087; Q53WT0; Q60635; Q61367; Q61427; Q63919;
 DT 01-JUN-1989 (Rel. 11, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN Name=Col1a1; Synonyms=Colal;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
 OC Muridae; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]
 RN NUCLEOTIDE SEQUENCE.
 RC STRAIN=FVB/N;
 RX MEDLINE=96033240; PubMed=8535610;
 RA Li S.W., Khillan J., Prockop D.J.;

RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
 of type I procollagen.";
 RL Matrix Biol. 14:593-595(1995).
 RN [2]
 RP NUCLEOTIDE SEQUENCE OF 1-185 AND 1030-1453.
 RX MEDLINE=94092741; PubMed=8268229; DOI=10.1016/0167-4781(93)90016-7;
 RA Fenton S.P., Lamande S.R., Hannagan M., Stacey A., Jaenisch R.,
 RA Bateman J.F.;
 RT "Genomic sequence of mouse COL1A1 encoding the collagen propeptides.";
 RL Biochim. Biophys. Acta 1216:469-474(1993).
 RN [3]
 RP NUCLEOTIDE SEQUENCE OF 1-942.
 RC STRAIN=C57BL/6; TISSUE=Liver;
 RX MEDLINE=94344105; PubMed=8065328;
 RA Rhodes K., Rippe R.A., Umezawa A., Nehls M., Brenner D.A., Breindl M.;
 RT "DNA methylation represses the murine alpha 1(I) collagen promoter by
 an indirect mechanism.";
 RL Mol. Cell. Biol. 14:5950-5960(1994).
 RN [4]
 RP NUCLEOTIDE SEQUENCE OF 518-1128
 RX MEDLINE=86137403; PubMed=3841523; DOI=10.1016/0378-1119(85)90329-4;
 RA French B.T., Lee W.-H., Maul G.G.;
 RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I) collagen
 protein.";
 RL Gene 39:311-312(1985).
 RN [5]
 RP NUCLEOTIDE SEQUENCE OF 735-1130.
 RX MEDLINE=83141374; PubMed=6298597;
 RA Monson J.M., Friedman J., McCarthy B.J.;
 RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
 evidence for a mouse B1 element within the gene.";
 RL Mol. Cell. Biol. 2:1362-1371(1982).
 RN [6]
 RP NUCLEOTIDE SEQUENCE OF 735-878 AND 1005-1058.
 RX MEDLINE=83157109; PubMed=6219867;
 RA Monson J.M., McCarthy B.J.;
 RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
 evidence for insertions or deletions in gene coding sequences.";
 RL DNA 1:59-69(1981).
 RN [7]
 RP NUCLEOTIDE SEQUENCE OF 1-25.
 RX MEDLINE=84170331; PubMed=6324198;
 RA Harbers K., Kuehn M., Delius H., Jaenisch R.;
 RT "Insertion of retrovirus into the first intron of alpha 1(I) collagen
 gene leads to embryonic lethal mutation in mice.";
 RL Proc. Natl. Acad. Sci. U.S.A. 81:1504-1508(1984).
 RN [8]
 RP NUCLEOTIDE SEQUENCE OF 1442-1453.
 RX MEDLINE=88124276; PubMed=3340560;
 RA Mooslehner K., Harbers K.;
 RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size of
 the 3'-untranslated region.";
 RL Nucleic Acids Res. 16:773-773(1988).
 RN [9]
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA] OF 1442-1453.
 RX PubMed=2054384;
 RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
 RT "Specific hybridization probes for mouse type I, II, III and IX
 collagen mRNAs.";
 RL Biochim. Biophys. Acta 1089:241-243(1991).
 CC -I- FUNCTION: Type I collagen is a member of group I collagen
 (fibrillar forming collagen).
 CC -I- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -I- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
 bones. In bones the fibrils are mineralized with calcium
 hydroxyapatite.
 CC -I- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -I- SIMILARITY: Belongs to the fibrillar collagen family.
 CC -I- SIMILARITY: Contains 1 WFC domain.
 CC -I- CAUTION: Ref.3 sequence differs from that shown due to erroneous
 gene model prediction.
 CC -----

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

```

-----
DR EMBL; U08020; AAB89912.1; -; mRNA.
DR EMBL; S67530; AAB29424.1; -; Genomic DNA.
DR EMBL; S67482; -; NOT ANNOTATED CDS; Genomic DNA.
DR EMBL; X54876; CAA38657.1; ALT SEQ; Genomic DNA.
DR EMBL; M14423; AAB37333.1; -; mRNA.
DR EMBL; M17491; AAB37334.1; -; Genomic DNA.
DR EMBL; K03036; AAB37332.1; -; Genomic DNA.
DR EMBL; K03029; AAB37332.1; JOINED; Genomic DNA.
DR EMBL; K03030; AAB37332.1; JOINED; Genomic DNA.
DR EMBL; K03031; AAB37332.1; JOINED; Genomic DNA.
DR EMBL; K03032; AAB37332.1; JOINED; Genomic DNA.
DR EMBL; K03033; AAB37332.1; JOINED; Genomic DNA.
DR EMBL; K03034; AAB37332.1; JOINED; Genomic DNA.
DR EMBL; K03035; AAB37332.1; JOINED; Genomic DNA.
DR EMBL; K01688; AAB37330.1; -; Genomic DNA.
DR EMBL; X05753; CAA29927.1; -; Genomic DNA.
DR EMBL; X15896; CAA33904.1; -; Genomic DNA.
DR EMBL; X57981; CAA41046.1; -; Genomic DNA.
DR PIR; S57243; S21626.
DR Ensembl; ENSMUSG0000001506; Mus musculus.
DR MGI; MGI:88467; Collal.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001007; VWFC_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 3.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWFC; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS0184; VWFC_2; 1.
KW Collagen; Extracellular matrix; Glycoprotein; Hydroxylation;
KW Pyrrolidone carboxylic acid; Repeat; Signal; Structural protein.
FT SIGNAL 1 22
FT PROPEP 23 151 N-terminal propeptide.
FT CHAIN 152 1207 Collagen alpha 1(I) chain.
FT PROPEP 1208 1453 C-terminal propeptide.
FT DOMAIN 29 87 VWFC.
FT REGION 152 167 Nonhelical region (N-terminal).
FT REGION 168 1181 Triple-helical region.
FT REGION 1182 1207 Nonhelical region (C-terminal).
FT MOTIF 734 736 Cell attachment site (Potential).
FT MOTIF 1082 1084 Cell attachment site (Potential).
FT MOD_RES 152 152 Pyrrolidone carboxylic acid (By similarity).
FT MOD_RES 160 160 Allysine (By similarity).
FT MOD_RES 254 254 5-hydroxylysine (By similarity).
FT MOD_RES 1153 1153 3-hydroxyproline (By similarity).
FT CARBOHYD 56 56 N-linked (GlcNAc...) (Potential).
FT CARBOHYD 254 254 O-linked (Gal...) (By similarity).
FT CARBOHYD 1354 1354 N-linked (GlcNAc...) (By similarity).
FT CONFLICT 81 81 G -> E (in Ref. 2 and 3).
FT CONFLICT 106 106 G -> D (in Ref. 2 and 3).
FT CONFLICT 1202 1202 D -> G (in Ref. 2).
FT CONFLICT 1219 1219 A -> E (in Ref. 2).
FT CONFLICT 1222 1222 A -> T (in Ref. 2).
FT CONFLICT 1335 1335 T -> A (in Ref. 2).
FT CONFLICT 1399 1400 RV -> TL (in Ref. 2).
FT CONFLICT 1450 1450 A -> V (in Ref. 8).
SQ SEQUENCE 1453 AA; 137943 MW; 3B802E53DF81808 CRC64;

Query Match 91.6%; Score 1218; DB 1; Length 1453;
Best Local Similarity 89.8%; Pred. No. 5.7e-102;
Matches 220; Conservative 12; Mismatches 13; Indels 0; Gaps 0;

```

```

QY 3 DANVVRDRLEVDVDTLKSLSQIENIRSPGSRKKNPARTCRDLKMCCHSDWKSGEYWDPN 62
DB 1209 DANVVRDRDLAVDTLKSLSQIENIRSPGSRKKNPARTCRDLKMCCHSDWKSGEYWDPN 1268

QY 63 QGCNLDIAKVCNMGTCVYPTQPSVAQKMWYISKNPDKRHHVWFGESMTDGFQFEYG 122
DB 1269 QGCNLDIAKVCNMGTCVYPTQPSVAQKMWYISKNPDKRHHVWFGESMTDGFQFEYG 1328

QY 123 GQGSDDPADVAIQITFLRLMSTEQSNITVHCNKNVAYMDQQTGNLKKALLKGSNEIEIR 182
DB 1329 SEGSDPTDVAIQITFLRLMSTEQSNITVHCNKNVAYMDQQTGNLKKALLKGSNEIEIR 1388

QY 183 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKTSRLPIIDVAPLDVGAPDOEFGFDVG 242
DB 1389 GEGNSRFTYSRVVDGCTSHTGAWGKTIVIEYKTKTSRLPIIDVAPLDVGAPDOEFGFDVG 1448

QY 243 PVCFL 247
DB 1449 PACFV 1453

RESULT 13
Q60785_MOUSE PRELIMINARY; PRT; 325 AA.
ID Q60785_MOUSE PRELIMINARY; PRT; 325 AA.
AC Q60785;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Alpha-1 type I procollagen (Fragment).
GN Names=Collal;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Lu Z, Laptev A, V. Prockop D, J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL; U03419; AAB03475.1; -; mRNA.
DR MGI; MGI:88467; Collal.
DR GO; GO:0005615; C:extracellular space; TAS.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
KW Collagen.
FT NON_TER 1 1
SQ SEQUENCE 325 AA; 35229 MW; A5A21E74DFDE3EF1 CRC64;

Query Match 90.7%; Score 1205; DB 2; Length 325;
Best Local Similarity 89.0%; Pred. No. 1.3e-101;
Matches 218; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

QY 3 DANVVRDRLEVDVDTLKSLSQIENIRSPGSRKKNPARTCRDLKMCCHSDWKSGEYWDPN 62
DB 81 DANVVRDRDLAVDTLKSLSQIENIRSPGSRKKNPARTCRDLKMCCHSDWKSGEYWDPN 140

QY 63 QGCNLDIAKVCNMGTCVYPTQPSVAQKMWYISKNPDKRHHVWFGESMTDGFQFEYG 122
DB 141 QGCNLDIAKVCNMGTCVYPTQPSVAQKMWYISKNPDKRHHVWFGESMTDGFQFEYG 200

QY 123 GQGSDDPADVAIQITFLRLMSTEQSNITVHCNKNVAYMDQQTGNLKKALLKGSNEIEIR 182
DB 201 SEGSDPTDVAIQITFLRLMSTEQSNITVHCNKNVAYMDQQTGNLKKALLKGSNEIEIR 260

QY 183 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKTSRLPIIDVAPLDVGAPDOEFGFDVG 242
DB 261 GEGNSRFTYSRVVDGCTSHTGAWGKTIVIEYKTKTSRLPIIDVAPLDVGAPDOEFGFDVG 320

```

```

Qy 243 PVCEL 247
Db 321 PXCFV 325

RESULT 14
QY1B4_CYNPY
ID QY1B4_CYNPY PRELIMINARY; PRT; 1450 AA.
AC QY1B4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Alpha 1 type I collagen.
OS Cynops pyrrhogaster (Japanese common newt).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.
OX NCBI_TaxID=8330;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Regenerate forelimbs;
RX MEDLINE=99407244; PubMed=10474166;
RX DOI=10.1002/(SICI)1097-0177(199909)216:1<59::AID-DVDY8>3.3.CO;2-2;
RA Ashihara K., Obara M., Yoshizato K.;
RT "Expression of genes of type I and type II collagen in the formation
RT and development of the blastema of regenerating newt limb.";
RL Dev. Dyn. 216:59-71(1999).
DR EMBL; AB015438; BAA36973.1; -; mRNA.
DR GO; GO:0005581; C:cytoplasm; IEA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; UNKNOWN_1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen.
SQ SEQUENCE 1450 AA; 137563 MW; ABF8A74841B87B7C CRC64;

Query Match 90.78; Score 1205; DB 2; Length 1450;
Best Local Similarity 87.8%; Pred. No. 8.9e-101;
Matches 215; Conservative 19; Mismatches 11; Indels 0; Gaps 0;

Qy 3 DANVVRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db 1206 DANVVRDLVDVDTTLKSLAQIENIRSPGTRKNPARTCRDLKMKCHSDWKSQDYWDPN 1265

Qy 63 QGCLDAIKVFCNNETGETCVYPTQPSVAQKNWYISKNPDKKHVWFGEWNTGQFQPEYG 122
Db 1266 QGCLDAIKVHCNNETGETCVYPSQASISQKNWYTSKNPREKHVWFGEWNTGQFQPEYG 1325

Qy 123 QGSDPADVAIQLTFLRLMTEASONITYHCKSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1326 GEGSDPADVNQLTFLRLMTEASONITYHCKSVAYMDQETGNLKKAVLLQGSNEIEIR 1385

Qy 183 AEGNSRFTYSVTVDGCTSHTGAMGKTVIEYKTKSSRLPIIDVAPLDVVGAPDOBFQPDVG 242
Db 1386 AEGNSRFTYGVTEGCTQHTGEGKTVIEYKTKSRLPIIDVAPMDVGTPDQFEGIDIG 1445

Qy 243 PVCEL 247
Db 1446 PVCEL 1450

RESULT 15
Q640B2_XENTR

```

```

ID Q640B2_XENTR PRELIMINARY; PRT; 1449 AA.
AC Q640B2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical LOC496414.
GN Name=LOC496414;
OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus; Silurana.
OX NCBI_TaxID=8364;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Whole body;
RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC TISSUE=Whole body;
RX Klein S., Gerhard D.S.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC082718; AAH82718.1; -; mRNA.
DR Ensembl; ENSXETG0000003374; Xenopus tropicalis.
DR GO; GO:0005581; C:collagen; IEA.
DR GO; GO:0005201; F:extracellular matrix structural constituent; IEA.
DR GO; GO:0006817; P:phosphate transport; IEA.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR00885; Fib_collagen_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; VWC; 1.
DR ProDom; PD000007; Clg_helix; 4.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWF_C_1; 1.
DR PROSITE; PS0184; VWF_C_2; 1.
KW Collagen; Hypothetical protein.
SQ SEQUENCE 1449 AA; 137539 MW; 456639B1687A3B4B CRC64;

Query Match 89.2%; Score 1185; DB 2; Length 1449;
Best Local Similarity 86.5%; Pred. No. 6e-99;
Matches 212; Conservative 21; Mismatches 12; Indels 0; Gaps 0;

Qy 3 DANVVRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db 1205 DANVVRDLVDVDTTLKSLKQIENIRSPGTRKNPARTCRDLKMKCHSDWKSQDYWDPN 1264

Qy 63 QGCLDAIKVFCNNETGETCVYPTQPSVAQKNWYISKNPDKKHVWFGEWNTGQFQPEYG 122

```

Db 1265 QGCILDAIKYICNMETGETCVYPTONSIAQKNWYTSKNPREKXHVWFGEAMSDGQFXYG 1324

Qy 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182

Db 1325 SEGSDPADVAIQLTFLRLMATEASQNIYHCKNSVAYMDQOTGNLKKALLQGSNEIEIR 1384

Qy 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDOESFGFDVG 242

Db 1385 AEGNSRFTYSVVEDGCTQHTGENGKTVIDYKTKTSRLPITDVAPMDIGAPDOEFGIDIG 1444

Qy 243 PVCFL 247

Db 1445 PVCfV 1449

Search completed: April 26, 2006, 16:19:22
Job time : 134.566 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:16:43 ; Search time 18.5914 Seconds
(without alignments)
1098.405 Million cell updates/sec

Title: US-10-677-877A-4

Perfect score: 1329

Sequence: 1 RSDANVVRDLEVDVTLKS.....LDVGAPDQEGFDVGPVCFCL 247

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pap:*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pap:*
- 3: /cgn2_6/ptodata/1/iaa/H COMB.pap:*
- 4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap:*
- 5: /cgn2_6/ptodata/1/iaa/RE COMB.pap:*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1317	99.1	250	2	US-09-029-348-14
2	1317	99.1	1461	2	US-09-585-887-9
3	1317	99.1	1461	2	US-09-289-578-9
4	1317	99.1	1464	2	US-09-331-347C-21
5	1299.5	97.8	1341	2	US-08-963-825-18
6	1299.5	97.8	1341	2	US-09-500-811-18
7	1299.5	97.8	1341	2	US-09-570-573-18
8	1299.5	97.8	1341	2	US-09-548-608-18
9	982.5	73.9	1418	2	US-08-963-825-20
10	982.5	73.9	1418	2	US-09-500-811-20
11	982.5	73.9	1418	2	US-09-570-573-20
12	982.5	73.9	1418	2	US-09-548-608-20
13	977.5	73.6	1418	2	US-09-010-999-1
14	967.5	72.8	1442	1	US-08-316-650-12
15	967.5	72.8	1442	4	PCT-US95-02251-12
16	861.5	64.8	537	2	US-09-029-348-4
17	861.5	64.8	626	2	US-09-029-348-2
18	858	64.6	1366	2	US-09-585-887-10
19	858	64.6	1366	2	US-09-289-578-10
20	858	64.6	1366	2	US-09-949-016-5882
21	853	64.2	534	2	US-09-029-348-5
22	853	64.2	535	2	US-09-029-348-1
23	853	64.2	623	2	US-09-029-348-3
24	853	64.2	1366	2	US-08-963-825-19
25	853	64.2	1366	2	US-09-500-811-19
26	853	64.2	1366	2	US-09-570-573-19
27	853	64.2	1366	2	US-09-548-608-19

28	508.5	38.3	1806	2	US-09-919-497-56	Sequence 56, Appl
29	433.5	32.6	1745	2	US-09-795-061-4	Sequence 4, Appl
30	433.5	32.6	1745	2	US-09-949-002-405	Sequence 405, App
31	433.5	32.6	1771	2	US-09-949-002-492	Sequence 492, App
32	416	31.3	1739	2	US-09-795-061-2	Sequence 2, Appl
33	387.5	29.2	122	2	US-10-104-047-2717	Sequence 2717, Ap
34	114	8.6	23	2	US-09-029-348-7	Sequence 7, Appl
35	114	8.6	23	2	US-09-585-887-1	Sequence 1, Appl
36	114	8.6	23	2	US-09-289-578-1	Sequence 1, Appl
37	95	7.1	244	1	US-08-553-497A-22	Sequence 22, Appl
38	90.5	6.8	503	2	US-08-700-651-6	Sequence 6, Appl
39	90.5	6.8	503	2	US-09-588-995A-66	Sequence 66, Appl
40	90	6.8	641	2	US-09-328-352-5990	Sequence 5990, Ap
41	89	6.7	244	1	US-08-553-497A-20	Sequence 20, Appl
42	89	6.7	246	1	US-08-553-497A-24	Sequence 24, Appl
43	87	6.5	413	2	US-09-107-433-4507	Sequence 4507, Ap
44	87	6.5	643	1	US-08-245-511-47	Sequence 47, Appl
45	87	6.5	643	1	US-08-600-993A-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1

US-09-029-348-14
; Sequence 14, Application US/09029348
; Patent No. 6171827
; GENERAL INFORMATION:
; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER
; TITLE OF INVENTION: NOVEL PROCOLLAGENS
; FILE REFERENCE: d087857PUS LISTING
; CURRENT APPLICATION NUMBER: US/09/029,348
; CURRENT FILING DATE: 1998-05-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 250
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE
; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS
US-09-029-348-14

Query Match	99.1%	Score 1317;	DB 2;	Length 250;
Best Local Similarity	99.6%	Pred. No. 7.6e-135;		
Matches 244;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	3	DANVVRDRDLEVDVTLKSLSQIENIRSPGSKNPARTCRDLKMCCHSDKSGEYWDPN	62	
Db	6	DANVVRDRDLEVDVTLKSLSQIENIRSPGSKNPARTCRDLKMCCHSDKSGEYWDPN	65	
Qy	63	QGNLDAIKVFCNMETGTCVPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG	122	
Db	66	QGNLDAIKVFCNMETGTCVPTQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG	125	
Qy	123	QGSDDPADVAIQITFLRLMSTEASQNTYHCNKSVAQMDQQTGNLKKALLKGSNEIEIR	182	
Db	126	QGSDDPADVAIQITFLRLMSTEASQNTYHCNKSVAQMDQQTGNLKKALLKGSNEIEIR	185	
Qy	183	AEGNSRFTSVTVGDCITSTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEGFDVG	242	
Db	186	AEGNSRFTSVTVGDCITSTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEGFDVG	245	
Qy	243	PVCFL 247		
Db	246	PVCFL 250		

RESULT 2

US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742

```

; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-585-887-9

Query Match          99.1%; Score 1317; DB 2; Length 1461;
Best Local Similarity 99.6%; Pred. No. 1.1e-133;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLEVDVTTLKSLSSQOIENIRSPGSRKKNPARTCRDLKMCHSDWKSGEYWDPN 62
Db 1217 DANVVRDRDLEVDVTTLKSLSSQOIENIRSPGSRKKNPARTCRDLKMCHSDWKSGEYWDPN 1276

QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGSMTDGFQFEYG 122
Db 1277 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGSMTDGFQFEYG 1336

QY 123 QGGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1337 QGGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1396

QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1397 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 1456

QY 243 PVCFL 247
Db 1457 PVCFL 1461

RESULT 4
US-09-331-347C-21
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Met
; TITLE OF INVENTION: obtaining Such and Their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331,347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-331-347C-21

Query Match          99.1%; Score 1317; DB 2; Length 1464;
Best Local Similarity 99.6%; Pred. No. 1.1e-133;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLEVDVTTLKSLSSQOIENIRSPGSRKKNPARTCRDLKMCHSDWKSGEYWDPN 62
Db 1220 DANVVRDRDLEVDVTTLKSLSSQOIENIRSPGSRKKNPARTCRDLKMCHSDWKSGEYWDPN 1279

QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGSMTDGFQFEYG 122
Db 1280 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGSMTDGFQFEYG 1339

QY 123 QGGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1340 QGGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1399

QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 1459

QY 243 PVCFL 247
Db 1460 PVCFL 1464

RESULT 5
```

```

US-09-289-578-9

Query Match          99.1%; Score 1317; DB 2; Length 1461;
Best Local Similarity 99.6%; Pred. No. 1.1e-133;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLEVDVTTLKSLSSQOIENIRSPGSRKKNPARTCRDLKMCHSDWKSGEYWDPN 62
Db 1217 DANVVRDRDLEVDVTTLKSLSSQOIENIRSPGSRKKNPARTCRDLKMCHSDWKSGEYWDPN 1276

QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGSMTDGFQFEYG 122
Db 1277 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGSMTDGFQFEYG 1336

QY 123 QGGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1337 QGGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1396

QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1397 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 1456

QY 243 PVCFL 247
Db 1457 PVCFL 1461

RESULT 4
US-09-331-347C-21
; Sequence 21, Application US/09331347C
; Patent No. 6617431
; GENERAL INFORMATION:
; APPLICANT: Meristem Therapeutics, S.A.
; APPLICANT: Meristem Therapeutics, S.A.
; TITLE OF INVENTION: Recombinant Collagens and Derived Proteins Produced by Plants, Met
; TITLE OF INVENTION: obtaining Such and Their Uses
; FILE REFERENCE: 1149-3
; CURRENT APPLICATION NUMBER: US/09/331,347C
; CURRENT FILING DATE: 1999-08-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-331-347C-21

Query Match          99.1%; Score 1317; DB 2; Length 1464;
Best Local Similarity 99.6%; Pred. No. 1.1e-133;
Matches 244; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLEVDVTTLKSLSSQOIENIRSPGSRKKNPARTCRDLKMCHSDWKSGEYWDPN 62
Db 1220 DANVVRDRDLEVDVTTLKSLSSQOIENIRSPGSRKKNPARTCRDLKMCHSDWKSGEYWDPN 1279

QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGSMTDGFQFEYG 122
Db 1280 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGSMTDGFQFEYG 1339

QY 123 QGGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1340 QGGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1399

QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKTSRLPIIDVAPLDVGAPDQEFQFDVG 1459

QY 243 PVCFL 247
Db 1460 PVCFL 1464

RESULT 5
```

US-08-963-825-18
; Sequence 18, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; US-08-963-825-18

Query Match 97.8%; Score 1299.5; DB 2; Length 1341;
Best Local Similarity 98.8%; Pred. No. 7.9e-132;
Matches 243; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 3 DANVVRDRDLEVDTTLSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 62
DB 1096 DANVVRDRDLEVDTTLSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 1155
QY 63 QGCNLDIAKVFNCNMTGETCVYPTQPSVAQKNWYISKNPDK-RHVVFGSMTDGFQFEY 121
DB 1156 QGCNLDIAKVFNCNMTGETCVYPTQPSVAQKNWYISKNPDK-RHVVFGSMTDGFQFEY 1215
QY 122 GGQSGDPADVAIQLTFLRLMSTASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 181
DB 1216 GGQSGDPADVAIQLTFLRLMSTASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 1275
QY 182 RAEGNSRFTYVTVGCTSHTGAWGKTVIEYKTKSSRLPIIDVAPLDVGAPQDFGFDV 241
DB 1276 RAEGNSRFTYVTVGCTSHTGAWGKTVIEYKTKSSRLPIIDVAPLDVGAPQDFGFDV 1335
QY 242 GPNVCF 247
DB 1336 GPNVCF 1341

RESULT 6
US-09-500-811-18
; Sequence 18, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
; US-09-500-811-18
Query Match 97.8%; Score 1299.5; DB 2; Length 1341;
Best Local Similarity 98.8%; Pred. No. 7.9e-132;
Matches 243; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
QY 3 DANVVRDRDLEVDTTLSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 62
DB 1096 DANVVRDRDLEVDTTLSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 1155
QY 63 QGCNLDIAKVFNCNMTGETCVYPTQPSVAQKNWYISKNPDK-RHVVFGSMTDGFQFEY 121
DB 1156 QGCNLDIAKVFNCNMTGETCVYPTQPSVAQKNWYISKNPDK-RHVVFGSMTDGFQFEY 1215
QY 122 GGQSGDPADVAIQLTFLRLMSTASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 181
DB 1216 GGQSGDPADVAIQLTFLRLMSTASQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 1275
QY 182 RAEGNSRFTYVTVGCTSHTGAWGKTVIEYKTKSSRLPIIDVAPLDVGAPQDFGFDV 241
DB 1276 RAEGNSRFTYVTVGCTSHTGAWGKTVIEYKTKSSRLPIIDVAPLDVGAPQDFGFDV 1335

```
QY 242 GPCFL 247
Db 1336 GPCFL 1341

RESULT 7
US-09-570-573-18
; Sequence 18, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-570-573-18

Query Match 97.8%; Score 1299.5; DB 2; Length 1341;
Best Local Similarity 98.8%; Pred. No. 7.9e-132;
Matches 243; Conservative 0; Mismatches 1; Gaps 1;

QY 3 DANVVRDRLEVDVTTKLSLQQIENIRSPGSRKNPARTCRDLKMHCHSDWKSGEYWDPN 62
Db 1096 DANVVRDRLEVDVTTKLSLQQIENIRSPGSRKNPARTCRDLKMHCHSDWKSGEYWDPN 1155
QY 63 QGCNLDKIVFCNMETGETCVYPTQPSVAQKNWYISKNPDK-RHWFGESEMTDGFQFEY 121
Db 1156 QGCNLDKIVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEY 1215
QY 122 GGQSDPADVAIQTLFLRLMSTEQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 181
Db 1216 GGQSDPADVAIQTLFLRLMSTEQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 1275
QY 182 RAEGNSRFTYSVTVDGCTSHTGAWGKTVEIKTKSRLLPIIDVAPLDVGPQDEFGFDV 241

QY 242 GPCFL 247
Db 1336 GPCFL 1341

RESULT 8
US-09-548-608-18
; Sequence 18, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1341 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (I)
US-09-548-608-18

Query Match 97.8%; Score 1299.5; DB 2; Length 1341;
Best Local Similarity 98.8%; Pred. No. 7.9e-132;
Matches 243; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 3 DANVVRDRLEVDVTTKLSLQQIENIRSPGSRKNPARTCRDLKMHCHSDWKSGEYWDPN 62
Db 1096 DANVVRDRLEVDVTTKLSLQQIENIRSPGSRKNPARTCRDLKMHCHSDWKSGEYWDPN 1155
QY 63 QGCNLDKIVFCNMETGETCVYPTQPSVAQKNWYISKNPDK-RHWFGESEMTDGFQFEY 121
Db 1156 QGCNLDKIVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEY 1215
QY 122 GGQSDPADVAIQTLFLRLMSTEQNIYTHCKNSVAYMDQQTGNLKKALLKGSNEIEI 181
```

Db 1216 GGGSDPADVAIQITFLRLMSTEAQNITTHCKNSVAYMDQOTGNLKKALLXGKSNEIBI 1275
Qy 182 RAEGNSRFTYVTDGCTSHGTGAWKTVIYKTKSSRLPIIDVAPLDVGAPDQBFV 241
Db 1276 RAEGNSRFTYVTDGCTSHGTGAWKTVIYKTKSSRLPIIDVAPLDVGAPDQBFV 1335
Qy 242 GPVCFL 247
Db 1336 GPVCFL 1341

RESULT 9
US-08-963-825-20
; Sequence 20, Application US/08963825
; Patent No. 6110689
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,825
; FILING DATE:
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/187,319
; FILING DATE: 21-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-08-963-825-20

Query Match 73.9%; Score 982.5; DB 2; Length 1418;
Best Local Similarity 69.1%; Pred. No. 2.7e-97;
Matches 172; Conservative 36; Mismatches 38; Indels 3; Gaps 2;

Qy 1 RSD--ANVVRDRDLEVDITLKSLSQQIENIRSPGSRKNPARTCRDLKMHCHSDWKSGEYW 58
Db 1171 RADQAAGLRQHDVAEDVATLKSLSNQIESIRSPGSRKNPARTCRDLKCHPEWKSQDYW 1230

Qy 59 IDPQGNLDAIKVFCNMTGETCVYTPQPSVAQKNYVISKNPDKKHVWFGSMTDGFQ 118
Db 1231 IDPQGNLDAIKVFCNMTGETCVYTPNPANVPKNWSSKS-KEKKHWFGETINGGFH 1289

Qy 119 FEYGGGSDPADVAIQITFLRLMSTEAQNITTHCKNSVAYMDQOTGNLKKALLXGKSNE 178
Db 1290 FSYGDDNLAPNTANVQMTFLRLSTEGSNITTHCKNSIAYLDEAAAGNLKALLIQSND 1349
Qy 179 IEIRAGNSRFTYVTDGCTSHGTGAWKTVIYKTKSSRLPIIDVAPLDVGAPDQBFV 238
Db 1350 VEIRAGNSRFTYVTDGCTSHGTGAWKTVIYKTKSSRLPIIDVAPLDVGAPDQBFV 1409
Qy 239 FDVGPVCFL 247
Db 1410 VDIGPVCFL 1418

RESULT 10
US-09-500-811-20
; Sequence 20, Application US/09500811
; Patent No. 6323314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-09-500-811-20

Query Match 73.9%; Score 982.5; DB 2; Length 1418;
Best Local Similarity 69.1%; Pred. No. 2.7e-97;
Matches 172; Conservative 36; Mismatches 38; Indels 3; Gaps 2;

Qy 1 RSD--ANVVRDRDLEVDITLKSLSQQIENIRSPGSRKNPARTCRDLKMHCHSDWKSGEYW 58
Db 1171 RADQAAGLRQHDVAEDVATLKSLSNQIESIRSPGSRKNPARTCRDLKCHPEWKSQDYW 1230

QY 59 IDPNQCNLDIAKVFNCMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 118
Db 1231 IDPNQCNLDIAKVFNCMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 118
QY 119 FEYGGQSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQOQTNLKKALLKGSNE 178
Db 1290 FSYGDDNLAPNTANVQMTFLRLMSTEQSNITTHCKNSVAYMDQOQTNLKKALLKGSND 1349
QY 179 IEIRAEGNSRFTYVTVGCTSHGTGKGTVEYKTKSRRLPIIDVAPLDVCGAPQOEFQ 238
Db 1350 VEIRAEGNSRFTYVTVGCTSHGTGKGTVEYKTKSRRLPIIDVAPLDVCGAPQOEFQ 238
QY 239 FDVGPCVCF 247
Db 1410 VDIGPVCFL 1418

RESULT 11
US-09-570-573-20
; Sequence 20, Application US/09570573
; Patent No. 6342361
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/570,573
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-09-570-573-20

Query Match 73.9%; Score 982.5; DB 2; Length 1418;
Best Local Similarity 69.1%; Pred. No. 2.7e-97;
Matches 172; Conservative 36; Mismatches 38; Indels 3; Gaps 2;
QY 1 RSD--ANVVRDRDLEVDTLKSLSQIENIRSPGSRKNPACTCRDLKCHDWKSGEYW 58

Db 1171 RADQAAGLRQDAEVDATLKSINNOIESIRSPGSRKNPACTCRDLKCHDWKSGDYW 1230
QY 59 IDPNQCNLDIAKVFNCMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 118
Db 1231 IDPNQCNLDIAKVFNCMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 118
QY 119 FEYGGQSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQOQTNLKKALLKGSNE 178
Db 1290 FSYGDDNLAPNTANVQMTFLRLMSTEQSNITTHCKNSVAYMDQOQTNLKKALLKGSND 1349
QY 179 IEIRAEGNSRFTYVTVGCTSHGTGKGTVEYKTKSRRLPIIDVAPLDVCGAPQOEFQ 238
Db 1350 VEIRAEGNSRFTYVTVGCTSHGTGKGTVEYKTKSRRLPIIDVAPLDVCGAPQOEFQ 238
QY 239 FDVGPCVCF 247
Db 1410 VDIGPVCFL 1418

RESULT 12
US-09-548-608-20
; Sequence 20, Application US/09548608
; Patent No. 6355442
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/548,608
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN -ALPHA 1 (II)
US-09-548-608-20

Query Match 73.9%; Score 982.5; DB 2; Length 1418;
Best Local Similarity 69.1%; Pred. No. 2.7e-97;

Matches 172; Conservative 36; Mismatches 38; Indels 3; Gaps 2;
Qy 1 RSD--ANVVRDRLEVDVTTLSLQQIENIRSPGSRKKNPARTCRDLKCHSDWKSGEYW 58
Db 1171 RADQAAGLRQDAEVDATLKSNNQIESIRSPGSRKKNPARTCRDLKCHPEWKSQDYW 1230
Qy 59 IDPNQGNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 118
Db 1231 IDPNQGTLDAMKVFVCMETGETCVYENPANVPKKNWSSKS-KEKCHIWFGETINGGPH 1289
Qy 119 FEYGGQSDPADVAIQLTFLRLMSTEQSONITTHCKNSVAYMDQOQTNLKKALLKGSNE 178
Db 1290 FSYGDDNLAPNTANVQMTFLRLSTEGSQNTTHCKNSIAYLDEAAGNLKALLIQSND 1349
Qy 179 IEIRAGNSRFTYSVTVDGCTSHTGAWGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEF 238
Db 1350 VEIRAGNSRFTYALKDGGCTKHTGKWKTVIEYRSQKTSRLPIIDVAPMDIGPEQEF 1409
Qy 239 FDVGPVCFL 247
Db 1410 VDIGPVCFL 1418

RESULT 13
US-09-010-999-1
; Sequence 1, Application US/09010999
; Patent No. 6132976
; GENERAL INFORMATION:
; APPLICANT: Poole, Anthony R.
; APPLICANT: Hollander, Anthony P.
; APPLICANT: Billingshurst, R. C.
; TITLE OF INVENTION: IMMUNOSAYS FOR THE MEASUREMENT OF
; NUMBER OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/010,999
; FILING DATE: 22-JAN-1998
; CLASSIFICATION: 4335
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/448,501
; FILING DATE: 17-JUL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/984,123
; FILING DATE: 04-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 032931/0212
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1418 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORGANISM: Human Type II Collagen

US-09-010-999-1
Query Match 73.8%; Score 977.5; DB 2; Length 1418;
Best Local Similarity 68.7%; Pred. No. 9.4e-97;
Matches 171; Conservative 36; Mismatches 39; Indels 3; Gaps 2;
Qy 1 RSD--ANVVRDRLEVDVTTLSLQQIENIRSPGSRKKNPARTCRDLKCHSDWKSGEYW 58
Db 1171 RADQAAGLRQDAEVDATLKSNNQIESIRSPGSRKKNPARTCRDLKCHPEWKSQDYW 1230
Qy 59 IDPNQGNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 118
Db 1231 IDPNQGTLDAMKVFVCMETGETCVYENPANVPKKNWSSKS-KEKCHIWFGETINGGPH 1289
Qy 119 FEYGGQSDPADVAIQLTFLRLMSTEQSONITTHCKNSVAYMDQOQTNLKKALLKGSNE 178
Db 1290 FSYGDDNLAPNTANVQMTFLRLSTEGSQNTTHCKNCIAYLDEAAGNLKALLIQSND 1349
Qy 179 IEIRAGNSRFTYSVTVDGCTSHTGAWGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEF 238
Db 1350 VEIRAGNSRFTYALKDGGCTKHTGKWKTVIEYRSQKTSRLPIIDVAPMDIGPEQEF 1409
Qy 239 FDVGPVCFL 247
Db 1410 VDIGPVCFL 1418
RESULT 14
US-08-316-650-12
; Sequence 12, Application US/08316650
; Patent No. 5942496
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Roessler, Blake J.
; APPLICANT: Goldstein, Steven A.
; APPLICANT: Lin, Wushan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS
; TITLE OF INVENTION: FOR STIMULATING BONE CELLS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC:008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-08-316-650-12

Query Match 72.8%; Score 967.5; DB 1; Length 1442;
Best Local Similarity 68.3%; Pred. No. 1.2e-95;
Matches 166; Conservative 39; Mismatches 37; Indels 1; Gaps 1;
QY 5 NVVRDRDLVDVDTLKSLSQOIEINIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDIPNOG 64
DB 1201 STLRQHDVVDVDTLKSLSQOIEINIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDIPNOG 1260
QY 65 CNLDAIKVFCNNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGEISMTDGFQFYGQ 124
DB 1261 CTLDAMKVFNNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGEISMTDGFQFYGQ 1319
QY 125 GSDPADVAIQTLFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNEIEIRAE 184
DB 1320 NLAPNTANVQMTFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNEIEIRAE 1379
QY 185 GNSRFTYSVTVDGCTSHGTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGPDPQDFGPDVGPV 244
DB 1380 GNSRFTYTALKDQCTKHTGKGTIVIEYRSQKTSRLPIIDVAPLDVGPDPQDFGPDVGPV 1439
QY 245 CFL 247
DB 1440 CFL 1442

RESULT 15

PCT-US95-02251-12
; Sequence 12, Application PC/TUS9502251
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
; TITLE OF INVENTION: CELLS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/02251
; FILING DATE: CONCURRENTLY HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,650
; FILING DATE: 30-SEP-1994
; CLASSIFICATION:
; APPLICATION NUMBER: US 08/199,780
; FILING DATE: 18-FEB-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UMIC009P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1442 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

; MOLECULE TYPE: peptide
PCT-US95-02251-12

Query Match 72.8%; Score 967.5; DB 4; Length 1442;
Best Local Similarity 68.3%; Pred. No. 1.2e-95;
Matches 166; Conservative 39; Mismatches 37; Indels 1; Gaps 1;
QY 5 NVVRDRDLVDVDTLKSLSQOIEINIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDIPNOG 64
DB 1201 STLRQHDVVDVDTLKSLSQOIEINIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDIPNOG 1260
QY 65 CNLDAIKVFCNNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGEISMTDGFQFYGQ 124
DB 1261 CTLDAMKVFNNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGEISMTDGFQFYGQ 1319
QY 125 GSDPADVAIQTLFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNEIEIRAE 184
DB 1320 NLAPNTANVQMTFLRLMSTEQSNITTHCKNSVAYMDQOTGNLKKALLKGSNEIEIRAE 1379
QY 185 GNSRFTYSVTVDGCTSHGTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGPDPQDFGPDVGPV 244
DB 1380 GNSRFTYTALKDQCTKHTGKGTIVIEYRSQKTSRLPIIDVAPLDVGPDPQDFGPDVGPV 1439
QY 245 CFL 247
DB 1440 CFL 1442

Search completed: April 26, 2006, 16:18:01
Job time : 19.5914 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:17:23 ; Search time 50.905 Seconds
(without alignments)
2027.381 Million cell updates/sec

Title: US-10-677-877A-4

Perfect score: 1329

Sequence: 1 RSDANVVRDRDLEVDITLKS.....LDVGAPDQEGFVGPVCF 247

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubaa/US09_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1329	100.0	247	5	US-10-677-877A-4
2	1329	100.0	635	5	US-10-677-877A-16
3	1329	100.0	755	5	US-10-677-877A-8
4	1324	99.6	502	5	US-10-677-877A-12
5	1320	99.3	258	4	US-10-139-794-81
6	1320	99.3	311	5	US-10-677-877A-2
7	1320	99.3	566	5	US-10-677-877A-10
8	1320	99.3	699	5	US-10-677-877A-14
9	1320	99.3	819	5	US-10-677-877A-6
10	1320	99.3	1464	3	US-09-918-715-261
11	1320	99.3	1464	4	US-10-060-036-159
12	1320	99.3	1464	4	US-10-171-311-36
13	1320	99.3	1464	4	US-10-149-352-2
14	1320	99.3	1464	4	US-10-177-293-65
15	1320	99.3	1464	4	US-10-301-822-28
16	1320	99.3	1464	4	US-10-734-564-79
17	1320	99.3	1464	4	US-10-474-794-261
18	1320	99.3	1464	5	US-10-723-860-2289
19	1320	99.3	1464	5	US-10-852-335A-157
20	1320	99.3	1464	5	US-10-979-159-261
21	1320	99.3	1464	5	US-10-287-436A-1187
22	1320	99.3	1464	5	US-10-287-436A-1187
23	1317	99.1	246	4	US-10-366-125-13
24	1317	99.1	250	5	US-10-632-847-16
25	1317	99.1	1461	4	US-10-468-091-25
26	1317	99.1	1464	4	US-10-216-705-21
27	1317	99.1	1464	4	US-10-764-425-114

28 1317 99.1 1464 4 US-10-357-851-1 Sequence 1, Appli
29 1317 99.1 1464 4 US-10-358-024-1 Sequence 1, Appli
30 1317 99.1 1464 4 US-10-788-792-150 Sequence 150, App
31 1313 98.8 449 3 US-09-925-301-884 Sequence 884, App
32 1313 98.8 1464 4 US-10-291-265-243 Sequence 243, App
33 1299.5 97.8 1341 4 US-10-058-124-18 Sequence 18, Appl
34 1292 97.2 420 4 US-10-468-091-24 Sequence 24, Appl
35 1290 97.1 1463 4 US-10-402-089-2 Sequence 2, Appli
36 1290 97.1 1463 4 US-10-402-072A-2 Sequence 2, Appli
37 1276 96.0 1449 4 US-10-402-089-8 Sequence 8, Appli
38 1276 96.0 1449 4 US-10-402-072A-8 Sequence 8, Appli
39 1267 95.3 441 4 US-10-106-698-6367 Sequence 6367, Ap
40 1218 91.6 1453 4 US-10-468-091-26 Sequence 26, Appl
41 1199 90.2 220 4 US-10-139-794-201 Sequence 201, App
42 1005.5 75.7 246 4 US-10-366-125-17 Sequence 17, Appl
43 982.5 73.9 1418 4 US-10-058-124-20 Sequence 20, Appl
44 982.5 73.9 1418 4 US-10-468-091-5 Sequence 5, Appli
45 982.5 73.9 1487 5 US-10-756-149-4739 Sequence 4739, Ap

ALIGNMENTS

RESULT 1

US-10-677-877A-4

; Sequence 4, Application US/10677877A

; Publication No. US20050202537A1

; GENERAL INFORMATION:

; APPLICANT: Liang, Peng; GenHunter Corporation for producing secreted trimeric

; TITLE OF INVENTION: Methods and composition for producing active fusion proteins

; FILE REFERENCE: 03-052-PL

; CURRENT APPLICATION NUMBER: US/10/677,877A

; CURRENT FILING DATE: 2003-10-02

; NUMBER OF SEQ ID NOS: 16

; SEQ ID NO 4

; LENGTH: 247

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-677-877A-4

Query Match 100.0%; Score 1329; DB 5; Length 247;
Best Local Similarity 100.0%; Pred. No. 8.9e-126; Indels 0; Gaps 0;
Matches 247; Conservative 0; Mismatches 0;

QY 1 RSDANVVRDRDLEVDITLKSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWD 60

DB 1 RSDANVVRDRDLEVDITLKSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWD 60

QY 61 PNQGCNLDIAIKVPCNMETGETCYPTQPSVAQKNWYISKPKDKRHWVFGESMTDGFQFE 120

DB 61 PNQGCNLDIAIKVPCNMETGETCYPTQPSVAQKNWYISKPKDKRHWVFGESMTDGFQFE 120

QY 121 YGQGGSDPADVAIQTLFLRLMSTEASQNTYHCNKSVAVMDDQQTGNLKALLKGSNEIE 180

DB 121 YGQGGSDPADVAIQTLFLRLMSTEASQNTYHCNKSVAVMDDQQTGNLKALLKGSNEIE 180

QY 181 IRAEGNSRFTYSVTVDGCTSHTGAWKTVIEYTKTSRLPIIDVAPLDVGAPDQEGF 240

DB 181 IRAEGNSRFTYSVTVDGCTSHTGAWKTVIEYTKTSRLPIIDVAPLDVGAPDQEGF 240

QY 241 VGPVCF 247

DB 241 VGPVCF 247

RESULT 2

US-10-677-877A-16

; Sequence 16, Application US/10677877A

; Publication No. US20050202537A1

; GENERAL INFORMATION:

; APPLICANT: Liang, Peng; GenHunter Corporation for producing secreted trimeric

; TITLE OF INVENTION: Methods and composition for producing secreted trimeric

; TITLE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 16
; LENGTH: 635
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-677-877A-16

Query Match 100.0%; Score 1329; DB 5; Length 635;
Best Local Similarity 100.0%; Pred. No. 3.3e-125;
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDANVVRDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKXCHSDWKSGBYWD 60
Db 389 RSDANVVRDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKXCHSDWKSGBYWD 448

Qy 61 PNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFE 120
Db 449 PNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFE 508

Qy 121 YGGGSDPADVAIQLTFLRLMSTEQSNITYHCKNSVAYMDQOTGNLKKALLKGSNEIE 180
Db 509 YGGGSDPADVAIQLTFLRLMSTEQSNITYHCKNSVAYMDQOTGNLKKALLKGSNEIE 568

Qy 181 IRAEGNSRFTYSVTVDCGTSHTGAWGKTVEYTKTSSRLPIIDVAPLDVGAPDQFGFD 240
Db 569 IRAEGNSRFTYSVTVDCGTSHTGAWGKTVEYTKTSSRLPIIDVAPLDVGAPDQFGFD 628

Qy 241 VGPVCFL 247
Db 629 VGPVCFL 635

RESULT 3
US-10-677-877A-8
; Sequence 8, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 8
; LENGTH: 755
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-677-877A-8

Query Match 100.0%; Score 1329; DB 5; Length 755;
Best Local Similarity 100.0%; Pred. No. 4.1e-125;
Matches 247; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RSDANVVRDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKXCHSDWKSGBYWD 60
Db 509 RSDANVVRDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKXCHSDWKSGBYWD 568

Qy 61 PNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFE 120
Db 569 PNQGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFE 628

Qy 121 YGGGSDPADVAIQLTFLRLMSTEQSNITYHCKNSVAYMDQOTGNLKKALLKGSNEIE 180
Db 629 YGGGSDPADVAIQLTFLRLMSTEQSNITYHCKNSVAYMDQOTGNLKKALLKGSNEIE 688

Qy 181 IRAEGNSRFTYSVTVDCGTSHTGAWGKTVEYTKTSSRLPIIDVAPLDVGAPDQFGFD 240
Db 689 IRAEGNSRFTYSVTVDCGTSHTGAWGKTVEYTKTSSRLPIIDVAPLDVGAPDQFGFD 748

Qy 241 VGPVCFL 247
Db 749 VGPVCFL 755

RESULT 4
US-10-677-877A-12
; Sequence 12, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677,877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 12
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-677-877A-12

Query Match 99.6%; Score 1324; DB 5; Length 502;
Best Local Similarity 100.0%; Pred. No. 7.6e-125;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 SDANVVRDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKXCHSDWKSGBYWD 61
Db 257 SDANVVRDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKXCHSDWKSGBYWD 316

Qy 62 NOGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFE 121
Db 317 NOGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFE 376

Qy 122 GGGSDPADVAIQLTFLRLMSTEQSNITYHCKNSVAYMDQOTGNLKKALLKGSNEIE 181
Db 377 GGGSDPADVAIQLTFLRLMSTEQSNITYHCKNSVAYMDQOTGNLKKALLKGSNEIE 436

Qy 182 RAEGNSRFTYSVTVDCGTSHTGAWGKTVEYTKTSSRLPIIDVAPLDVGAPDQFGFD 241
Db 437 RAEGNSRFTYSVTVDCGTSHTGAWGKTVEYTKTSSRLPIIDVAPLDVGAPDQFGFD 496

Qy 242 GPVCF 247
Db 497 GPVCF 502

RESULT 5
US-10-139-794-81
; Sequence 81, Application US/10139794
; Publication No. US20030232421A1
; GENERAL INFORMATION:
; APPLICANT: HYBERGENICS, LYNX THERAPEUTICS INC.
; APPLICANT: Pierre Legrain, Simon Whiteside, Jen-I Mao, Irina Khrebtukova, Shujun Luo
; TITLE OF INVENTION: Protein-Protein Interactions In Adipocyte Cells (3)
; FILE REFERENCE: B4883A
; CURRENT APPLICATION NUMBER: US/10/139,794
; CURRENT FILING DATE: 2002-05-06
; PRIOR APPLICATION NUMBER: US 60/288,885
; PRIOR FILING DATE: 2001-05-04
; NUMBER OF SEQ ID NOS: 2930
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 81
; LENGTH: 258
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: Translation of SEQ ID NO:80
US-10-139-794-81

Query Match 99.3%; Score 1320; DB 4; Length 258;

Best Local Similarity 100.0%; Pred. No. 7.7e-125;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
DB 14 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 73
QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEYG 122
DB 74 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEYG 133
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
DB 134 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 193
QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGPADQEFQFDVG 242
DB 194 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGPADQEFQFDVG 253
QY 243 PVCFL 247
DB 254 PVCFL 258

RESULT 6
US-10-677-877A-2
; Sequence 2, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677.877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 2
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-677-877A-2

Query Match 99.3%; Score 1320; DB 5; Length 311;
Best Local Similarity 100.0%; Pred. No. 1e-124;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
DB 67 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 126
QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEYG 122
DB 127 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEYG 186
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
DB 187 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 246
QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGPADQEFQFDVG 242
DB 247 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGPADQEFQFDVG 306
QY 243 PVCFL 247
DB 307 PVCFL 311

RESULT 7
US-10-677-877A-10
; Sequence 10, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation

; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677.877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 10
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-677-877A-10

Query Match 99.3%; Score 1320; DB 5; Length 566;
Best Local Similarity 100.0%; Pred. No. 2.3e-124;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
DB 322 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 381
QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEYG 122
DB 382 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEYG 441
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
DB 442 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 501
QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGPADQEFQFDVG 242
DB 502 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGPADQEFQFDVG 561
QY 243 PVCFL 247
DB 562 PVCFL 566

RESULT 8
US-10-677-877A-14
; Sequence 14, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; FILE REFERENCE: 03-052-PL
; CURRENT APPLICATION NUMBER: US/10/677.877A
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 14
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-677-877A-14

Query Match 99.3%; Score 1320; DB 5; Length 699;
Best Local Similarity 100.0%; Pred. No. 3e-124;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
DB 455 DANVVRDRDLVDVDTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 514
QY 63 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEYG 122
DB 515 QGCNLDIAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESEMTDGFQFEYG 574
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
DB 575 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 634
QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGPADQEFQFDVG 242

```
Db 635 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 694
QY 243 PVCFL 247
Db 695 PVCFL 699

RESULT 9
US-10-677-877A-6
; Sequence 6, Application US/10677877A
; Publication No. US20050202537A1
; GENERAL INFORMATION:
; APPLICANT: Liang, Peng; GenHunter Corporation
; TITLE OF INVENTION: Methods and composition for producing secreted trimeric
; TITLE OF INVENTION: receptor analogs and biologically active fusion proteins
; FILE REFERENCE: 03-052-FL
; CURRENT FILING DATE: 2003-10-02
; NUMBER OF SEQ ID NOS: 16
; SEQ ID NO 6
; LENGTH: 819
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-677-877A-6
Query Match 99.3%; Score 1320; DB 5; Length 819;
Best Local Similarity 100.0%; Pred. No. 3.8e-124; Indels 0; Gaps 0;
Matches 245; Conservative 0; Mismatches 0;

QY 3 DANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMCCHSWKSGEYWDPN 62
Db 575 DANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMCCHSWKSGEYWDPN 634
QY 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 635 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 694
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 695 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 754
QY 183 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 242
Db 755 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 814
QY 243 PVCFL 247
Db 815 PVCFL 819

RESULT 10
US-09-918-715-261
; Sequence 261, Application US/09918715
; Publication No. US20030017157A1
; GENERAL INFORMATION:
; APPLICANT: Brad St. Croix
; APPLICANT: Bert Vogelstein
; APPLICANT: Kenneth Kinzler
; TITLE OF INVENTION: ENDOTHELIAL CELL EXPRESSION PATTERNS
; FILE REFERENCE: 1107.00134
; CURRENT FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: 60/222,599
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 60/224,360
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/282,850
; PRIOR FILING DATE: 2000-04-11
; NUMBER OF SEQ ID NOS: 358
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 261
; LENGTH: 1464
; TYPE: PRT

; ORGANISM: Homo sapiens
US-09-918-715-261
Query Match 99.3%; Score 1320; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 8.4e-124;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMCCHSWKSGEYWDPN 62
Db 1220 DANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMCCHSWKSGEYWDPN 1279
QY 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 1280 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 1339
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1340 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1399
QY 183 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 1459
QY 243 PVCFL 247
```

```
; ORGANISM: Homo sapiens
US-09-918-715-261
Query Match 99.3%; Score 1320; DB 3; Length 1464;
Best Local Similarity 100.0%; Pred. No. 8.4e-124; Indels 0; Gaps 0;
Matches 245; Conservative 0; Mismatches 0;

QY 3 DANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMCCHSWKSGEYWDPN 62
Db 1220 DANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMCCHSWKSGEYWDPN 1279
QY 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 1280 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 1339
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1340 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1399
QY 183 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 1459
QY 243 PVCFL 247
Db 1460 PVCFL 1464

RESULT 11
US-10-060-036-159
; Sequence 159, Application US/10060036
; Publication No. US20030073144A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Persing, David H.
; APPLICANT: Hepler, William T.
; APPLICANT: Jiang, Yugu
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF PANCREATIC CANCER
; FILE REFERENCE: 210121.566
; CURRENT FILING DATE: 2002-01-30
; NUMBER OF SEQ ID NOS: 4560
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 159
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-060-036-159
Query Match 99.3%; Score 1320; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 8.4e-124;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMCCHSWKSGEYWDPN 62
Db 1220 DANVVRDRDLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMCCHSWKSGEYWDPN 1279
QY 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 1280 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 1339
QY 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1340 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1399
QY 183 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGAWGKTIVIEYKTKSSRLPIIDVAPLDVGAPDQDFGFDVG 1459
QY 243 PVCFL 247
```

```
Db      1460 PFCFL 1464
|||||
RESULT 12
US-10-171-311-36
; Sequence 36, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoerssch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF CERVICAL CANCER
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171.311
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-171-311-36

Query Match      99.3%; Score 1320; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 8.4e-124;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 DANVVRDRDLEVDTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db      1220 DANVVRDRDLEVDTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1279

Qy      63 QGCNLDKIVFCNMETGTCVYPTQPSVAQKNWYISKNPDKKRVHWFGEWMTDGFQFEYG 122
Db      1280 QGCNLDKIVFCNMETGTCVYPTQPSVAQKNWYISKNPDKKRVHWFGEWMTDGFQFEYG 1339

Qy      123 QGSDPADVAIQTLTFLRLMSTEQNTYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
Db      1340 QGSDPADVAIQTLTFLRLMSTEQNTYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1399

Qy      183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQBFQPDVG 242
Db      1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQBFQPDVG 1459

Qy      243 PFCFL 247
Db      1460 PFCFL 1464

RESULT 13
US-10-149-352-2
; Sequence 2, Application US/10149352
; Publication No. US20030105050A1
; GENERAL INFORMATION:
; APPLICANT: Berl, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/10/149.352
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
```

```
; PRIOR FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-149-352-2

Query Match      99.3%; Score 1320; DB 4; Length 1464;
Best Local Similarity 100.0%; Pred. No. 8.4e-124;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      3 DANVVRDRDLEVDTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db      1220 DANVVRDRDLEVDTTLSLSQQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1279

Qy      63 QGCNLDKIVFCNMETGTCVYPTQPSVAQKNWYISKNPDKKRVHWFGEWMTDGFQFEYG 122
Db      1280 QGCNLDKIVFCNMETGTCVYPTQPSVAQKNWYISKNPDKKRVHWFGEWMTDGFQFEYG 1339

Qy      123 QGSDPADVAIQTLTFLRLMSTEQNTYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
Db      1340 QGSDPADVAIQTLTFLRLMSTEQNTYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1399

Qy      183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQBFQPDVG 242
Db      1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQBFQPDVG 1459

Qy      243 PFCFL 247
Db      1460 PFCFL 1464

RESULT 14
US-10-177-293-65
; Sequence 65, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoerssch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,595
```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 26, 2006, 16:18:13 ; Search time 9.73835 Seconds
(without alignments)
1153.484 Million cell updates/sec

Title: US-10-677-877A-4
Perfect score: 1329
Sequence: 1 RSDANVVRDRDLEVDITLKS.....LDVGAPQERGFDPVCFVL 247

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 232119 seqs, 45477862 residues

Total number of hits satisfying chosen parameters: 232119

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /SID55/ptodata/2/pubpaa/US08 NEW PUB pep.*
 - 2: /SID55/ptodata/2/pubpaa/US06 NEW PUB pep.*
 - 3: /SID55/ptodata/2/pubpaa/US07 NEW PUB pep.*
 - 4: /SID55/ptodata/2/pubpaa/PCT NEW PUB pep.*
 - 5: /SID55/ptodata/2/pubpaa/US09 NEW PUB pep.*
 - 6: /SID55/ptodata/2/pubpaa/US10 NEW PUB pep.*
 - 7: /SID55/ptodata/2/pubpaa/US11 NEW PUB pep.*
 - 8: /SID55/ptodata/2/pubpaa/US60 NEW PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1320	99.3	1464	6	US-10-501-035-331
2	1320	99.3	1464	7	US-11-186-284-28
3	1320	99.3	1464	7	US-11-021-603-2
4	1313	98.8	1464	7	US-11-000-463-243
5	1313	98.8	1467	6	US-10-821-234-1096
6	1238	93.2	1453	6	US-10-784-004-344
7	1238	93.2	1453	6	US-10-784-004-434
8	1238	93.2	1453	6	US-10-784-004-918
9	1238	93.2	1453	6	US-10-784-004-950
10	982.5	73.9	1487	7	US-11-202-057-3
11	982.5	73.9	1487	7	US-11-202-057-5
12	982.5	73.9	1487	7	US-11-202-057-7
13	912.5	68.7	218	7	US-11-202-057-28
14	861.5	64.8	1166	6	US-10-821-234-964
15	861.5	64.8	1466	7	US-11-186-284-33
16	858	64.6	1366	6	US-10-821-234-1431
17	858	64.6	1366	7	US-11-186-284-31
18	853.5	64.2	636	6	US-10-784-004-428
19	755	56.8	1496	6	US-11-186-284-35
20	511.5	38.5	1874	6	US-10-821-234-1182
21	508.5	38.3	1806	7	US-11-051-720-1446
22	508.5	38.3	1806	7	US-11-051-720-1447
23	502.5	37.8	1767	6	US-10-995-561-911
24	502.5	37.8	1767	6	US-10-995-561-914
25	502.5	37.8	1806	6	US-10-995-561-912

RESULT 2

26	502.5	37.8	1806	6	US-10-995-561-915	Sequence 915, App
27	502.5	37.8	1818	6	US-10-995-561-910	Sequence 910, App
28	502.5	37.8	1818	6	US-10-995-561-913	Sequence 913, App
29	399.5	30.1	1736	7	US-11-124-368A-329	Sequence 329, App
30	387.5	29.2	122	7	US-11-072-512-2717	Sequence 2717, App
31	348.5	26.2	832	6	US-10-204-639-41	Sequence 41, Appl
32	100	7.5	248	7	US-11-054-515-1381	Sequence 1381, App
33	100	7.5	248	7	US-11-266-444-1381	Sequence 1381, App
34	88	6.6	254	7	US-11-054-515-1001	Sequence 1001, App
35	88	6.6	254	7	US-11-054-515-1334	Sequence 1334, App
36	88	6.6	254	7	US-11-266-444-1001	Sequence 1001, App
37	88	6.6	254	7	US-11-266-444-1334	Sequence 1334, App
38	87.5	6.6	251	7	US-11-054-515-1688	Sequence 1688, App
39	87.5	6.6	251	7	US-11-266-444-1688	Sequence 1688, App
40	87	6.5	254	7	US-11-054-515-873	Sequence 873, App
41	87	6.5	254	7	US-11-054-515-1088	Sequence 1088, App
42	87	6.5	254	7	US-11-266-444-873	Sequence 873, App
43	87	6.5	254	7	US-11-266-444-1088	Sequence 1088, App
44	86	6.5	254	7	US-11-054-515-1087	Sequence 1087, App
45	86	6.5	254	7	US-11-266-444-1087	Sequence 1087, App

ALIGNMENTS

RESULT 1

- US-10-501-035-331
; Sequence 331, Application US/10501035
; Publication No. US20060046249A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF POLYNUCLEOTIDES AND POLYPEPTIDE FOR PREDICTING
; TITLE OF INVENTION: ACTIVITY OF COMPOUNDS THAT INTERACT WITH PROTEIN TYROSINE KINASE
; FILE REFERENCE: D0185 PCT
; CURRENT APPLICATION NUMBER: US/10/501,035
; CURRENT FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: US 60/350,061
; PRIOR FILING DATE: 2002-01-18
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 331
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-501-035-331

Query Match	99.3%	Score 1320;	DB 6;	Length 1464;
Best Local Similarity	100.0%	Pred. No. 2.5e-116;		
Matches 245;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	3	DANVVRDRDLEVDITLKSLSQQIENIRSPGSRKNPARTCRDLKMHSDWKSGEYWDPN	62	
Db	1220	DANVVRDRDLEVDITLKSLSQQIENIRSPGSRKNPARTCRDLKMHSDWKSGEYWDPN	1279	
Qy	63	QGCNLDAIKVFCNMTGETCVYTPQPSVAQKNYISKNPDKRHWFGESMTDGFQPEYG	122	
Db	1280	QGCNLDAIKVFCNMTGETCVYTPQPSVAQKNYISKNPDKRHWFGESMTDGFQPEYG	1339	
Qy	123	GGQSDPADVAIQLTFLRLMSTEASQNTTYHCKNSVAMDOQTGNLKKALLKGSNEIEIR	182	
Db	1340	GGQSDPADVAIQLTFLRLMSTEASQNTTYHCKNSVAMDOQTGNLKKALLKGSNEIEIR	1399	
Qy	183	AEGRSFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQGFQPDVG	242	
Db	1400	AEGRSFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQGFQPDVG	1459	

243 PFCFL 247
1460 PFCFL 1464

```
US-11-186-284-28
; Sequence 28, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: WPM01-029P2NM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-28

Query Match          99.3%; Score 1320; DB 7; Length 1464;
Best Local Similarity 100.0%; Pred. No. 2.5e-116;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVDRDRLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db 1220 DANVDRDRLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1279

QY 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHVWFGESMTDGFQFEYG 122
Db 1280 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHVWFGESMTDGFQFEYG 1339

QY 123 QGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1340 QGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1399

QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 1459

QY 243 PFCFL 247
Db 1460 PFCFL 1464

RESULT 3
US-11-021-603-2
; Sequence 2, Application US/11021603
; Publication No. US20060003954A1
; GENERAL INFORMATION:
; APPLICANT: Beri, Rajinder
; TITLE OF INVENTION: ANTISENSE OLIGONUCLEOTIDES
; FILE REFERENCE: 06275-254US1
; CURRENT APPLICATION NUMBER: US/11/021,603
; CURRENT FILING DATE: 2004-12-21
; PRIOR APPLICATION NUMBER: US/10/149,352
; PRIOR FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: PCT/GB00/04741
; PRIOR FILING DATE: 2000-12-12
```

```
; PRIOR APPLICATION NUMBER: GB 9929487.8
; PRIOR FILING DATE: 1999-12-15
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 4.0
; SEQ ID NO 2
; LENGTH: 1464
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-021-603-2

Query Match          99.3%; Score 1320; DB 7; Length 1464;
Best Local Similarity 100.0%; Pred. No. 2.5e-116;
Matches 245; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 DANVDRDRLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db 1220 DANVDRDRLEVDVTTLKSLSQIENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1279

QY 63 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHVWFGESMTDGFQFEYG 122
Db 1280 QGCNLDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHVWFGESMTDGFQFEYG 1339

QY 123 QGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 182
Db 1340 QGSDPADVAIQLTFLRLMSTEASQNIYTHCKNSVAYMDQOTGNLKKALLKGSNEIEIR 1399

QY 183 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGANGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 1459

QY 243 PFCFL 247
Db 1460 PFCFL 1464

RESULT 4
US-11-000-463-243
; Sequence 243, Application US/11000463
; Publication No. US20050266423A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y Tom
; APPLICANT: Liu, Chenchua
; APPLICANT: Asundi, Vinod
; APPLICANT: Chen, Rui-hong
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Zhou, Ping
; APPLICANT: Cao, Yi-Cheng
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 785CIP4CN
; CURRENT APPLICATION NUMBER: US/11/000,463
; CURRENT FILING DATE: 2004-11-29
; PRIOR APPLICATION NUMBER: 10/291,265
; PRIOR FILING DATE: 2002-11-08
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: 09/922,279
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 09/617,746
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 09/631,451
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 09/633,870
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 944
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 243
; LENGTH: 1464
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-000-463-243

Query Match      98.8%; Score 1313; DB 7; Length 1464;
Best Local Similarity 99.2%; Pred. No. 1.2e-115;
Matches 243; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DANVVRDRDLEVDTTLKSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 62
Db 1220 DANVVRDRDLEVDTTLKSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 1279
Qy 63 QGCNLDIAIKVFCNMETGETCVYTPQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 122
Db 1280 QGCNLDIAIKVFCNMETGETCVYTPQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 1339
Qy 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
Db 1340 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1399
Qy 183 AEGNSRFTYSVTVDGCTSHTGAWGKTVEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1400 AEGNSRFTYSVTVDGCTSHTGAWGKTVEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 1459
Qy 243 PFCFL 247
Db 1460 PFCFL 1464

RESULT 5
US-10-821-234-1096
; Sequence 1096, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt SEQ_genes Version 1.0
; SEQ ID NO 1096
; LENGTH: 1467
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1096

Query Match      98.8%; Score 1313; DB 6; Length 1467;
Best Local Similarity 99.2%; Pred. No. 1.2e-115;
Matches 243; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DANVVRDRDLEVDTTLKSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 62
Db 1223 DANVVRDRDLEVDTTLKSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 1282
Qy 63 QGCNLDIAIKVFCNMETGETCVYTPQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 122
Db 1283 QGCNLDIAIKVFCNMETGETCVYTPQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 1342
Qy 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
Db 1343 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1402
Qy 183 AEGNSRFTYSVTVDGCTSHTGAWGKTVEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1403 AEGNSRFTYSVTVDGCTSHTGAWGKTVEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 1462
Qy 243 PFCFL 247

Db 1463 PFCFL 1467

RESULT 6
US-10-784-004-344
; Sequence 344, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idex
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 344
; LENGTH: 1453
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-344

Query Match      93.2%; Score 1238; DB 6; Length 1453;
Best Local Similarity 91.0%; Pred. No. 1.4e-108;
Matches 223; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Qy 3 DANVVRDRDLEVDTTLKSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 62
Db 1209 DANVVRDRDLEVDTTLKSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 1268
Qy 63 QGCNLDIAIKVFCNMETGETCVYTPQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 122
Db 1269 QGCNLDIAIKVFCNMETGETCVYTPQPSVAQKNWYISKNPDKRHWVFGESMTDGFQFEYG 1328
Qy 123 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 182
Db 1329 GQGSDDPADVAIQLTFLRLMSTEASQNIYHCKNSVAYMDQQTGNLKKALLKGSNEIEIR 1388
Qy 183 AEGNSRFTYSVTVDGCTSHTGAWGKTVEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 242
Db 1389 AEGNSRFTYSVTVDGCTSHTGAWGKTVEYKTKSSRLPIIDVAPLDVGAPDQEFQFDVG 1448
Qy 243 PFCFL 247
Db 1449 PFCFL 1453

RESULT 7
US-10-784-004-434
; Sequence 434, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idex
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 434
; LENGTH: 1453
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-434

Query Match      93.2%; Score 1238; DB 6; Length 1453;
Best Local Similarity 91.0%; Pred. No. 1.4e-108;
Matches 223; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Qy 3 DANVVRDRDLEVDTTLKSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 62
Db 1209 DANVVRDRDLEVDTTLKSLSQQIENIRSPGSRKNPARTCRDLKMCHSDWKSGEYWDPN 1268
```

QY 63 QGCNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 1269 QGCNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 1328
QY 123 QGSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 182
Db 1329 SEGSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 1388
QY 183 AEGNSRFTYVTDGCTSHGTGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 242
Db 1389 GEGNSRFTYVTDGCTSHGTGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 1448
QY 243 PVCFL 247
Db 1449 PACFV 1453

RESULT 8
US-10-784-004-918
; Sequence 918, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 918
; LENGTH: 1453
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-918

Query Match 93.2%; Score 1238; DB 6; Length 1453;
Best Local Similarity 91.0%; Pred. No. 1.4e-108;
Matches 223; Conservative 13; Mismatches 9; Indels 0; Gaps 0;
QY 3 DANVDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db 1209 DANVDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1268
QY 63 QGCNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 1269 QGCNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 1328
QY 123 QGSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 182
Db 1329 SEGSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 1388
QY 183 AEGNSRFTYVTDGCTSHGTGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 242
Db 1389 GEGNSRFTYVTDGCTSHGTGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 1448
QY 243 PVCFL 247
Db 1449 PACFV 1453

RESULT 9
US-10-784-004-950
; Sequence 950, Application US/10784004
; Publication No. US20060084066A1
; GENERAL INFORMATION:
; APPLICANT: Biogen Idec
; TITLE OF INVENTION: Surrogate Markers of Pain
; FILE REFERENCE: 08201.6029-00000
; CURRENT APPLICATION NUMBER: US/10/784,004
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 1251
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 950

; LENGTH: 1453
; TYPE: PRT
; ORGANISM: rat
US-10-784-004-950
Query Match 93.2%; Score 1238; DB 6; Length 1453;
Best Local Similarity 91.0%; Pred. No. 1.4e-108;
Matches 223; Conservative 13; Mismatches 9; Indels 0; Gaps 0;
QY 3 DANVDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 62
Db 1209 DANVDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPN 1268
QY 63 QGCNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 122
Db 1269 QGCNDAIKVFCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQFEYG 1328
QY 123 QGSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 182
Db 1329 SEGSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQGTGNLKKALLKGSNEIEIR 1388
QY 183 AEGNSRFTYVTDGCTSHGTGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 242
Db 1389 GEGNSRFTYVTDGCTSHGTGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQFVG 1448
QY 243 PVCFL 247
Db 1449 PACFV 1453

RESULT 10
US-11-202-057-3
; Sequence 3, Application US/11202057
; Publication No. US20060051794A1
; GENERAL INFORMATION:
; APPLICANT: TSAI, SHIH-FENG
; APPLICANT: LIU, YU-PEN
; APPLICANT: CHEN, WEI-MING
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO COL2A1 GENE
; TITLE OF INVENTION: MUTATIONS AND OSTEOECROSIS
; FILE REFERENCE: 08842.0011-00000
; CURRENT APPLICATION NUMBER: US/11/202,057
; CURRENT FILING DATE: 2005-08-12
; PRIOR APPLICATION NUMBER: 60/602,319
; PRIOR FILING DATE: 2004-08-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 3
; LENGTH: 1487
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-202-057-3

Query Match 73.9%; Score 982.5; DB 7; Length 1487;
Best Local Similarity 69.1%; Pred. No. 2.1e-84;
Matches 172; Conservative 36; Mismatches 38; Indels 3; Gaps 2;
QY 1 RSD--ANVDRDLVDVDTTLKSLSQOENIRSPGSRKNPARTCRDLKMKCHSDWKSGEY 58
Db 1240 RADQAAGLRQHDAYDATLKSNNQIESIRSPGSRKNPARTCRDLKMKCHSDWKSGEY 1299
QY 59 IDPNOCNLDIAKVFNCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 118
Db 1300 IDPNOCNLDIAKVFNCNMETGETCVYPTQPSVAQKNWYISKNPDKRHHVWFGESMTDGFQ 1358
QY 119 FEYGGGSDPADVAIQLTFLRLMSTEQSNITTHCKNSVAYMDQGTGNLKKALLKGSNE 178
Db 1359 FSYGDDNLAFTANVQWTLRLSLSTEGSQNITTHCKNSVAYMDQGTGNLKKALLKGSNE 1418
QY 179 IEIRAEGNSRFTYVTDGCTSHGTGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ 238
Db 1419 VEIRAEGNSRFTYVTDGCTSHGTGKTVIEYKTKSSRLPIIDVAPLDVGAPDQEFQ 1478

```
QY 239 FDVGPVCFL 247
Db 1479 VDGPVCFL 1487

RESULT 11
US-11-202-057-5
; Sequence 5, Application US/11202057
; Publication No. US20060051794A1
; GENERAL INFORMATION:
; APPLICANT: TSAI, SHIH-FENG
; APPLICANT: LIU, YU-FEN
; APPLICANT: CHEN, WEI-MING
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO COL2A1 GENE
; TITLE OF INVENTION: MUTATIONS AND OSTEOECROSIS
; FILE REFERENCE: 08842.0011-00000
; CURRENT APPLICATION NUMBER: US/11/202.057
; PRIOR FILING DATE: 2005-08-12
; PRIOR APPLICATION NUMBER: 60/602,319
; PRIOR FILING DATE: 2004-08-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 5
; LENGTH: 1487
; TYPE: PR1
; ORGANISM: Homo sapiens
US-11-202-057-5

Query Match 73.9%; Score 982.5; DB 7; Length 1487;
Best Local Similarity 69.1%; Pred. No. 2.1e-84;
Matches 172; Conservative 36; Mismatches 38; Indels 3; Gaps 2;

QY 1 RSD--ANVVRDRDLEVDVTTLSLSQOIENTRSPEGRKPNPARTCRDLKCHSDWKSGEYW 58
Db 1240 RAOQAAGLGHDAEVDATLKSNNQIESIRSPGRKPNPARTCRDLKCHPEWKSQDYW 1299

QY 59 IDNQCGLDAIKVFCNMETGETCVPTQPSVAKNWIYISKNPKDKRHVWFGEISMTDGFQ 118
Db 1300 IDNQCGLDAMKVFNCNMETGETCVYENPANVPKKNWSSKS-KEKKHIWFGETINGGFH 1358

QY 119 FEYGGQSDPADVAIQTLFLRLMSTASQNTYHCKNSVAYMDQOTGNLKKALLKGSNE 178
Db 1359 FSYGDDNLAPNTANVQMTFLRLSTEGSQNTYHCKNSIAYLDEAAGNLKALLIQSND 1418

QY 179 IEIRAEGNSRFTYSVTVDGCTSHTGAWKTVIEYKTKSSRLPLIIDVAPLDVGAPDQEF 238
Db 1419 VEIRAEGNSRFTYALKDGTCTHGTGKWKTVIEYRSQKTSRLPLIIDAPMDIGGPEQEF 1478

QY 239 FDVGPVCFL 247
Db 1479 VDGPVCFL 1487

RESULT 12
US-11-202-057-7
; Sequence 7, Application US/11202057
; Publication No. US20060051794A1
; GENERAL INFORMATION:
; APPLICANT: TSAI, SHIH-FENG
; APPLICANT: LIU, YU-FEN
; APPLICANT: CHEN, WEI-MING
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO COL2A1 GENE
; TITLE OF INVENTION: MUTATIONS AND OSTEOECROSIS
; FILE REFERENCE: 08842.0011-00000
; CURRENT APPLICATION NUMBER: US/11/202.057
; CURRENT FILING DATE: 2005-08-12
; PRIOR APPLICATION NUMBER: 60/602,319
; PRIOR FILING DATE: 2004-08-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 7
; LENGTH: 1487
; TYPE: PR1
US-11-202-057-7

Query Match 68.7%; Score 912.5; DB 7; Length 218;
Best Local Similarity 70.8%; Pred. No. 8.3e-79;
Matches 155; Conservative 32; Mismatches 31; Indels 1; Gaps 1;

QY 28 IRSPEGRKPNPARTCRDLKCHSDWKSGEYWIDPNQCNLDIAKVFNCNMETGETCVYPTQ 87
Db 1 IRSPEGRKPNPARTCRDLKCHPEWKSQDYWDPNQCTLDAMKVFNCNMETGETCVYPNP 60

QY 88 PSVAQKNWIYISKNPKDKRHVWFGEISMTDGFQFYSYGGSDPADVAIQTLFLRLMSTASQ 147
Db 61 ANVPKKNWSSKS-KEKKHIWFGETINGGFHPSYGGDNLAPNTANVQMTFLRLSTEGSQ 119

QY 148 NIYTHCKNSVAYMDQOTGNLKKALLKGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWK 207
Db 120 NIYTHCKNSIAYLDEAAGNLKALLIQSNDVIEIRAEGNSRFTYALKDGTCTHGTGKWK 179

QY 208 TVIEYKTKSSRLPLIIDVAPLDVGAPDQEFQFVDGPVCF 246
Db 180 TVIEYRSQKTSRLPLIIDAPMDIGGPEQEFVDGPVCF 218

RESULT 14
```

```
; ORGANISM: Homo sapiens
US-11-202-057-7

Query Match 73.9%; Score 982.5; DB 7; Length 1487;
Best Local Similarity 69.1%; Pred. No. 2.1e-84;
Matches 172; Conservative 36; Mismatches 38; Indels 3; Gaps 2;

QY 1 RSD--ANVVRDRDLEVDVTTLSLSQOIENTRSPEGRKPNPARTCRDLKCHPEWKSQDYW 58
Db 1240 RAOQAAGLGHDAEVDATLKSNNQIESIRSPGRKPNPARTCRDLKCHPEWKSQDYW 1299

QY 59 IDNQCGLDAIKVFCNMETGETCVPTQPSVAKNWIYISKNPKDKRHVWFGEISMTDGFQ 118
Db 1300 IDNQCGLDAMKVFNCNMETGETCVYENPANVPKKNWSSKS-KEKKHIWFGETINGGFH 1358

QY 119 FEYGGQSDPADVAIQTLFLRLMSTASQNTYHCKNSVAYMDQOTGNLKKALLKGSNE 178
Db 1359 FSYGDDNLAPNTANVQMTFLRLSTEGSQNTYHCKNSIAYLDEAAGNLKALLIQSND 1418

QY 179 IEIRAEGNSRFTYSVTVDGCTSHTGAWKTVIEYKTKSSRLPLIIDVAPLDVGAPDQEF 238
Db 1419 VEIRAEGNSRFTYALKDGTCTHGTGKWKTVIEYRSQKTSRLPLIIDAPMDIGGPEQEF 1478

QY 239 FDVGPVCFL 247
Db 1479 VDGPVCFL 1487

RESULT 13
US-11-202-057-28
; Sequence 28, Application US/11202057
; Publication No. US20060051794A1
; GENERAL INFORMATION:
; APPLICANT: TSAI, SHIH-FENG
; APPLICANT: LIU, YU-FEN
; APPLICANT: CHEN, WEI-MING
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO COL2A1 GENE
; TITLE OF INVENTION: MUTATIONS AND OSTEOECROSIS
; FILE REFERENCE: 08842.0011-00000
; CURRENT APPLICATION NUMBER: US/11/202.057
; CURRENT FILING DATE: 2005-08-12
; PRIOR APPLICATION NUMBER: 60/602,319
; PRIOR FILING DATE: 2004-08-18
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 28
; LENGTH: 218
; TYPE: PR1
; ORGANISM: Homo sapiens
US-11-202-057-28

Query Match 68.7%; Score 912.5; DB 7; Length 218;
Best Local Similarity 70.8%; Pred. No. 8.3e-79;
Matches 155; Conservative 32; Mismatches 31; Indels 1; Gaps 1;

QY 28 IRSPEGRKPNPARTCRDLKCHSDWKSGEYWIDPNQCNLDIAKVFNCNMETGETCVYPTQ 87
Db 1 IRSPEGRKPNPARTCRDLKCHPEWKSQDYWDPNQCTLDAMKVFNCNMETGETCVYPNP 60

QY 88 PSVAQKNWIYISKNPKDKRHVWFGEISMTDGFQFYSYGGSDPADVAIQTLFLRLMSTASQ 147
Db 61 ANVPKKNWSSKS-KEKKHIWFGETINGGFHPSYGGDNLAPNTANVQMTFLRLSTEGSQ 119

QY 148 NIYTHCKNSVAYMDQOTGNLKKALLKGSNEIEIRAEGNSRFTYSVTVDGCTSHTGAWK 207
Db 120 NIYTHCKNSIAYLDEAAGNLKALLIQSNDVIEIRAEGNSRFTYALKDGTCTHGTGKWK 179

QY 208 TVIEYKTKSSRLPLIIDVAPLDVGAPDQEFQFVDGPVCF 246
Db 180 TVIEYRSQKTSRLPLIIDAPMDIGGPEQEFVDGPVCF 218

RESULT 14
```

```
US-10-821-234-964
; Sequence 964, Application US/10821234
; Publication No. US2005025511A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; CURRENT APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 964
; LENGTH: 1166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-964

Query Match      64.8%; Score 861.5; DB 6; Length 1166;
Best Local Similarity 66.5%; Pred. No. 4.3e-73;
Matches 157; Conservative 32; Mismatches 44; Indels 3; Gaps 3;

Qy 13 EVDTTLKSLSQOENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPNQGCNLDAIKV 72
Db 933 EIMTSLKSVNGQIESLSPGSRKNPARNCRDILKFKHPELKSGEYWDPNQGCNLDAIKV 992
Qy 73 FCNMETGETCVYPTQPSVAQKNWYISKNPDKRHVWFGEISMTDGFQFEYGGQSGDPADVA 132
Db 993 FCNMETGETCISANPLNVPKHHW-TDSSAEKKHVVFGESMDGGFQFSYGNP-ELPEDVL 1050
Qy 133 -IQLTFLRLMSTEAQNITYHCKNSVAYMDQOQGNLKKALLKSGSNEIEIRAGNSRFTY 191
Db 1051 DVQLAFLRLLSRASQNIYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTY 1110
Qy 192 SVTVDGCTSHGTGAWGKTIVLEYKTKSSRLPIIDVAPLDVCGAPDQERGFVDGVPVCEL 247
Db 1111 TVLEDGCTKHTGWSKTVFEYTRKAVRLPIVDIAPYDIGPDPQERGFVDGVPVCEL 1166
```

```
RESULT 15
US-11-186-284-33
; Sequence 33, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MEM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
```

```
; LENGTH: 1466
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-33

Query Match      64.8%; Score 861.5; DB 7; Length 1466;
Best Local Similarity 66.5%; Pred. No. 5.7e-73;
Matches 157; Conservative 32; Mismatches 44; Indels 3; Gaps 3;

Qy 13 EVDTTLKSLSQOENIRSPGSRKNPARTCRDLKMKCHSDWKSGEYWDPNQGCNLDAIKV 72
Db 1233 EIMTSLKSVNGQIESLSPGSRKNPARNCRDILKFKHPELKSGEYWDPNQGCNLDAIKV 1292
Qy 73 FCNMETGETCVYPTQPSVAQKNWYISKNPDKRHVWFGEISMTDGFQFEYGGQSGDPADVA 132
Db 1293 FCNMETGETCISANPLNVPKHHW-TDSSAEKKHVVFGESMDGGFQFSYGNP-ELPEDVL 1350
Qy 133 -IQLTFLRLMSTEAQNITYHCKNSVAYMDQOQGNLKKALLKSGSNEIEIRAGNSRFTY 191
Db 1351 DVQLAFLRLLSRASQNIYHCKNSIAYMDQASGNVKKALKLMGSNEGEFKAEGNSKFTY 1410
Qy 192 SVTVDGCTSHGTGAWGKTIVLEYKTKSSRLPIIDVAPLDVCGAPDQERGFVDGVPVCEL 247
Db 1411 TVLEDGCTKHTGWSKTVFEYTRKAVRLPIVDIAPYDIGPDPQERGFVDGVPVCEL 1466

Search completed: April 26, 2006, 16:20:40
Job time : 10.7384 secs
```